

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 27, 2001, 08:16:52 ; Search time 2286.42 Seconds
(without alignments)
49.243 Million cell updates/sec

Title: US-09-380-826A-2
Perfect score: 22
Sequence: 1 ttttgatcacaaagtattgata 22

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1118133 seqs, 2558875100 residues

Total number of hits satisfying chosen parameters: 2236266

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba1.*
- 2: gb_ba2.*
- 3: gb_om.*
- 4: gb_ov.*
- 5: gb_ph.*
- 6: gb_pl1.*
- 7: gb_pl2.*
- 8: gb_pri.*
- 9: gb_pr2.*
- 10: gb_pr3.*
- 11: gb_ro.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: em_fun.*
- 15: em_hum1.*
- 16: em_hum2.*
- 17: em_in.*
- 18: em_om.*
- 19: em_or.*
- 20: em_ov.*
- 21: em_pat.*
- 22: em_ph.*
- 23: em_pl.*
- 24: em_ro.*
- 25: em_sts.*
- 26: em_sy.*
- 27: em_un.*
- 28: em_v1.*
- 29: gb_ba3.*
- 30: gb_in1.*
- 31: gb_in2.*
- 32: gb_in3.*
- 33: gb_pl3.*
- 34: gb_pr4.*
- 35: em_ba1.*
- 36: em_ba2.*
- 37: em_htg1.*
- 38: em_htg2.*
- 39: em_htg3.*
- 40: em_htg4.*
- 41: em_htg5.*
- 42: em_htg6.*
- 43: em_htg7.*

- 44: em_htg8.*
- 45: em_htg9.*
- 46: em_htg10.*
- 47: em_hum3.*
- 48: em_hum4.*
- 49: em_hum5.*
- 50: em_hum6.*
- 51: gb_pr5.*
- 52: gb_pr6.*
- 53: gb_pr7.*
- 54: gb_htg1.*
- 55: gb_htg2.*
- 56: gb_htg3.*
- 57: gb_htg4.*
- 58: gb_htg5.*
- 59: gb_htg6.*
- 60: gb_htg7.*
- 61: gb_htg8.*
- 62: gb_htg9.*
- 63: gb_htg10.*
- 64: gb_htg11.*
- 65: gb_htg12.*
- 66: gb_htg13.*
- 67: gb_htg14.*
- 68: gb_htg15.*
- 69: gb_htg16.*
- 70: gb_htg17.*
- 71: gb_htg18.*
- 72: gb_htg19.*
- 73: gb_htg20.*
- 74: gb_htg21.*
- 75: gb_htg22.*
- 76: gb_htg23.*
- 77: gb_sts1.*
- 78: gb_sts2.*
- 79: gb_v1.*
- 80: gb_v12.*
- 81: gb_pat1.*
- 82: gb_pat2.*
- 83: em_htg0.*
- 84: gb_htg24.*
- 85: gb_pr8.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	22	100.0	1481	2	LFU60594
2	20.4	92.7	354	1	AB007014
3	18.8	85.5	283	1	AF003953
4	18.8	85.5	288	2	LI094974
5	18.8	85.5	288	2	LI094975
6	18.8	85.5	288	2	LI094976
7	18.8	85.5	288	2	LI094977
8	18.8	85.5	288	2	LI094978
9	18.8	85.5	288	2	LI094979
10	18.8	85.5	353	1	AB007012
11	18.8	85.5	1012	7	ATHRS2X
12	18.8	85.5	72590	6	AC003000
13	18.8	85.5	109723	72	AF165146
14	18.8	85.5	200799	6	AF137379
15	18.8	85.5	200799	6	AF137379
c 16	18.4	83.6	1812	31	AF220067
c 17	18.4	83.6	27365	58	AC015209
18	18.4	83.6	185087	54	AC007475
19	18.4	83.6	262731	30	AE003823
20	17.8	80.9	1874	79	EBBPC44NCP
c 21	17.8	80.9	31214	33	SPCC830
					U60594 Leptosira
					AB007014 Spirochae
					AF003953 Leptospir
					U94974 Leptosira
					U94975 Leptosira
					U94976 Leptosira
					U94977 Leptosira
					U94978 Leptosira
					U94979 Leptosira
					AB007012 Spirochae
					M98336 Arabidopsis
					AC003000 Arabidops
					AF165146 Homo sapi
					AF137379 Nephrocel
					AF137379 Nephrocel
					AF220067 Drosophil
					AC015209 Drosophil
					AC007475 Drosophil
					AE003823 Drosophil
					L48441 Echinocloa
					AL109850 S.pombe c

Wed Mar 28 14:03:14 2001

AB009052 Arabidops
AC026327 Homo sapi
AC026320 Homo sapi
AC055742 Homo sapi
AC003114 Arabidops
AB008265 Arabidops
AC068144 Homo sapi
AC021015 Homo sapi
AC023155 Homo sapi
AF175672 Unculture
AF271231 Albunio ca
AF018567 Unidentif
D49367 Lithospermu
Z73136 S.cerevisia
U40832 Strongyloce
U41016 Caenorhabdi
U41545 Caenorhabdi
AL049803 Arabidops
AC083790 Homo sapi
AC036233 Homo sapi
AP002082 Homo sapi
AL078469 Arabidops
AL078470 Arabidops
Z98052 Human DNA s

AB009052
80.9 85992 6 AB009052
80.9 91894 65 AC026327
80.9 174766 65 AC026320
80.9 270889 67 AC055742
79.1 59261 33 T12M4
79.1 81662 6 AB008265
79.1 109512 68 AC068144
79.1 149241 61 AC021015
79.1 187847 62 AC023155
78.2 462 1 AF175672
78.2 835 7 AF271231
78.2 1083 1 AF018567
78.2 2213 33 LEP4CCOALB
78.2 3608 33 SCYLL031C
78.2 5596 32 SP0A0832
78.2 34496 31 CELR11G1
78.2 38655 31 CELC02F12
78.2 49311 84 F26K10
78.2 62478 71 AC083790
78.2 67858 66 AC036233
78.2 68040 76 AP002082
78.2 97714 7 ATF25024
78.2 104679 7 ATF19B15
78.2 112467 52 HS055B13

ALIGNMENTS

RESULT 1
LFU60594 1481 bp DNA BCT 10-SEP-1998
LOCUS Leptospiira fainei 16S ribosomal RNA gene, partial sequence.
AC0594
U60594.1 GI:1408219
Leptospiira fainei.
Bacteria: Spirochaetales; Leptospiiraceae; Leptospiira.
1 (bases 1 to 1481)
Perolat,P., Chappel,R.J., Adler,B., Baranton,G., Bulach,D.M.,
Billinghurst,M.L., Letocart,M., Merien,F. and Serrano,M.S.
Leptospiira fainei sp. nov., isolated from pigs in Australia
Int. J. Syst. Bacteriol. 48 Pt 3, 851-858 (1998)
98404550
2 (bases 1 to 1481)
Adler,B., Chappel,R.J., Baranton,G., Bulach,D.M.,
Billinghurst,M.L., Letocart,M., Merien,F., Serrano,M.S. and
Perolat,P.
Direct Submission
Submitted (12-JUN-1996) Microbiology, Monash University, Wellington
Rd., Clayton, VIC 3168, Australia
Location/Qualifiers
1..1481
/organism="Leptospiira fainei"
/strain="Hurstbridge"
/db_xref="taxon:48782"
1..>1481
/product="16S ribosomal RNA"
391 a 335 c 439 g 314 t 2 others

Query Match 100.0%; Score 22; DB 2; Length 1481;
Best Local Similarity 100.0%; Pred. No. 3.9;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 tggtagatcacaagattgata 22
|||||
Db 155 TGTGGATCACAAGATTGATA 176

RESULT 2

AB007014 354 bp DNA BCT 13-OCT-1997
LOCUS Spirochaeta sp. 16S rRNA gene, partial sequence.
AC007014
AB007014.1 GI:2516255
16S ribosomal RNA.
Spirochaeta sp. (sub-species: Freshwater obligate oligotroph,
strain: FO-95) DNA.
Spirochaeta sp.
Bacteria; Spirochaetales; Spirochaetaeae; Spirochaeta.
1 (bases 1 to 354)
Shin,M.-S.
Direct Submission
Submitted (05-SEP-1997) to the DDBJ/EMBL/GenBank databases. Mi-Sun
Shin, Laboratory of Marine Molecular Microbiology, Faculty of
Agriculture, Kyoto University; Oiwakecho, Kitashirakawa, Sakyo-ku,
Kyoto, Kyoto 306-01, Japan (E-mail: misun@kais.kyoto-u.ac.jp,
Tel: 075-753-6224, Fax: 075-753-6226)
2 (sites)
Shin,M., Yoshinaga,I., Uchida,A. and Ishida,Y.
Phylogenetic analysis by 16S rRNA gene sequencing of Obligate
Oligotrophs isolated from the northern basin of Lake Biwa
(Mesotrophic Lake)
Unpublished (1997)
Location/Qualifiers
1..354
/organism="Spirochaeta sp."
/strain="FO-95"
/sub-species="Freshwater obligate oligotroph"
/db_xref="taxon:28185"
1..>354
/product="16S rRNA"
91 a 79 c 111 g 73 t
BASE COUNT
ORIGIN
Query Match 92.7%; Score 20.4; DB 1; Length 354;
Best Local Similarity 95.5%; Pred. No. 25;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Oy 1 tggtagatcacaagattgata 22
|||||
Db 133 TGTGGATCACAAGATTGATA 154

RESULT 3
AF003953 283 bp DNA BCT 30-MAY-1998
LOCUS Leptospiira interrogans strain 48/95 16S ribosomal RNA gene, partial
sequence.
AC003953
AF003953.1 GI:3169306
Leptospiira interrogans.
Bacteria; Spirochaetales; Leptospiiraceae; Leptospiira.
1 (bases 1 to 283)
Woo,T.H.S., Patel,B.K.C., Cinco,M., Smythe,L.D., Symonds,M.,
Norris,M. and Dohnt,M.
Identification of Leptonema by real-time homogeneous assay of rapid
cycle PCR product
Unpublished
2 (bases 1 to 283)
Woo,T.H.S., Patel,B.K.C., Cinco,M., Smythe,L.D., Symonds,M.,
Norris,M. and Dohnt,M.
Direct Submission
Submitted (14-MAY-1997) School of Science, Griffith University,
Brisbane, QLD 4111, Australia
Location/Qualifiers
1..283
/organism="Leptospiira interrogans"
/strain="48/95"
/db_xref="taxon:173"

```

rRNA
<1..>283
/product="16S rRNA"
74 a 60 c 88 g 61 t
BASE COUNT
ORIGIN

Query Match      85.5%; Score 18.8; DB 1; Length 283;
Best Local Similarity 90.9%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 tgttgatcacaaagatttgata 22
| | | | | | | | | | | | | | | |
Db 136 TATTGGATCACAGGATTTGATA 157

RESULT 4
LIU94974
LOCUS      LIU94974      288 bp      DNA      BCT      01-JAN-1998
DEFINITION Leptospira inadai 16S ribosomal RNA gene, partial sequence.
ACCESSION  U94974
VERSION     U94974.1 GI:2735446
KEYWORDS   Leptospira inadai.
SOURCE     Leptospira inadai.
ORGANISM   Bacteria; Spirochaetales; Leptospiaceae; Leptospira.
REFERENCE  1 (bases 1 to 288)
AUTHORS    Woo, T.H.S., Smythe, L.D., Symonds, M., Norris, M., Dohnt, M.,
            Brenner, D.J. and Patel, B.K.C.
TITLE      Identification of Leptospira inadai by continuously monitoring
            fluorescence during rapid cycle PCR
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 288)
AUTHORS    Woo, T.H.S., Smythe, L.D., Symonds, M., Norris, M., Dohnt, M.,
            Brenner, D.J. and Patel, B.K.C.
TITLE      Direct Submission
JOURNAL    Submitted (24-MAR-1997) School of Science, Griffith University,
            Brisbane, QLD 4111, Australia
FEATURES   source
            Location/Qualifiers
            1..288
            /organism="Leptospira inadai"
            /strain="79/95"
            /db_xref="taxon:29506"
            <1..>288
            /product="16S ribosomal RNA"
            77 a 61 c 88 g 62 t
BASE COUNT
ORIGIN

Query Match      85.5%; Score 18.8; DB 2; Length 288;
Best Local Similarity 90.9%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 tgttgatcacaaagatttgata 22
| | | | | | | | | | | | | | | |
Db 141 TATTGGATCACAGGATTTGATA 162

RESULT 5
LIU94975
LOCUS      LIU94975      288 bp      DNA      BCT      01-JAN-1998
DEFINITION Leptospira inadai 16S ribosomal RNA gene, partial sequence.
ACCESSION  U94975
VERSION     U94975.1 GI:2735447
KEYWORDS   Leptospira inadai.
SOURCE     Leptospira inadai.
ORGANISM   Bacteria; Spirochaetales; Leptospiaceae; Leptospira.
REFERENCE  1 (bases 1 to 288)
AUTHORS    Woo, T.H.S., Smythe, L.D., Symonds, M., Norris, M., Dohnt, M.,
            Brenner, D.J. and Patel, B.K.C.
TITLE      Identification of Leptospira inadai by continuously monitoring
            fluorescence during rapid cycle PCR
JOURNAL    Unpublished

```

```

REFERENCE
AUTHORS    Woo, T.H.S., Smythe, L.D., Symonds, M., Norris, M., Dohnt, M.,
            Brenner, D.J. and Patel, B.K.C.
TITLE      Direct Submission
JOURNAL    Submitted (24-MAR-1997) School of Science, Griffith University,
            Brisbane, QLD 4111, Australia
FEATURES   source
            Location/Qualifiers
            1..288
            /organism="Leptospira inadai"
            /strain="68/94"
            /db_xref="taxon:29506"
            <1..>288
            /product="16S ribosomal RNA"
            77 a 61 c 88 g 62 t
BASE COUNT
ORIGIN

Query Match      85.5%; Score 18.8; DB 2; Length 288;
Best Local Similarity 90.9%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 tgttgatcacaaagatttgata 22
| | | | | | | | | | | | | | | |
Db 141 TATTGGATCACAGGATTTGATA 162

RESULT 6
LIU94976
LOCUS      LIU94976      288 bp      DNA      BCT      01-JAN-1998
DEFINITION Leptospira inadai 16S ribosomal RNA gene, partial sequence.
ACCESSION  U94976
VERSION     U94976.1 GI:2735448
KEYWORDS   Leptospira inadai.
SOURCE     Leptospira inadai.
ORGANISM   Bacteria; Spirochaetales; Leptospiaceae; Leptospira.
REFERENCE  1 (bases 1 to 288)
AUTHORS    Woo, T.H.S., Smythe, L.D., Symonds, M., Norris, M., Dohnt, M.,
            Brenner, D.J. and Patel, B.K.C.
TITLE      Identification of Leptospira inadai by continuously monitoring
            fluorescence during rapid cycle PCR
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 288)
AUTHORS    Woo, T.H.S., Smythe, L.D., Symonds, M., Norris, M., Dohnt, M.,
            Brenner, D.J. and Patel, B.K.C.
TITLE      Direct Submission
JOURNAL    Submitted (24-MAR-1997) School of Science, Griffith University,
            Brisbane, QLD 4111, Australia
FEATURES   source
            Location/Qualifiers
            1..288
            /organism="Leptospira inadai"
            /strain="346/95"
            /db_xref="taxon:29506"
            <1..>288
            /product="16S ribosomal RNA"
            77 a 61 c 88 g 62 t
BASE COUNT
ORIGIN

Query Match      85.5%; Score 18.8; DB 2; Length 288;
Best Local Similarity 90.9%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 tgttgatcacaaagatttgata 22
| | | | | | | | | | | | | | | |
Db 141 TATTGGATCACAGGATTTGATA 162

RESULT 7
LIU94977
LOCUS      LIU94977      288 bp      DNA      BCT      01-JAN-1998
DEFINITION Leptospira inadai 16S ribosomal RNA gene, partial sequence.
ACCESSION  U94977

```

Wed Mar 28 14:03:14 2001

us-09-380-826a-2.rge

U94977.1 GI:2735449
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Leptospiira inadai.
 Bacteria; Spirochaetales; Leptospiiraceae; Leptospiira.
 REFERENCE
 1 (bases 1 to 288)
 Woo, T.H.S., Smythe, L.D., Symonds, M., Norris, M., Dohnt, M.,
 Brenner, D.J. and Patel, B.K.C.
 Identification of Leptospiira inadai by continuously monitoring
 fluorescence during rapid cycle PCR
 TITLE
 JOURNAL
 REFERENCE
 2 (bases 1 to 288)
 Woo, T.H.S., Smythe, L.D., Symonds, M., Norris, M., Dohnt, M.,
 Brenner, D.J. and Patel, B.K.C.
 Direct Submission
 TITLE
 JOURNAL
 Submitted (24-MAR-1997) School of Science, Griffith University,
 Brisbane, QLD 4111, Australia
 FEATURES
 Location/Qualifiers
 source
 1..288
 /organism="Leptospiira inadai"
 /strain="268/95"
 /db_xref="taxon:29506"
 <1..>288
 /product="16S ribosomal RNA"
 77 a 61 c 88 g 62 t
 BASE COUNT
 ORIGIN

Query Match 85.5%; Score 18.8; DB 2; Length 288;
 Best Local Similarity 90.9%; Pred. No. 1.4e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 tgttgatcacagaatttgata 22
 | ||||| ||||| ||||| |||||
 Db 141 TATTGGATCACAGGATTGATA 162

RESULT 8
 LIU94978 288 bp DNA BCT 01-JAN-1998
 LOCUS
 DEFINITION
 Leptospiira inadai 16S ribosomal RNA gene, partial sequence.
 ACCESSION
 U94978
 VERSION
 U94978.1 GI:2735450
 KEYWORDS
 SOURCE
 Leptospiira inadai.
 ORGANISM
 Bacteria; Spirochaetales; Leptospiiraceae; Leptospiira.
 REFERENCE
 1 (bases 1 to 288)
 Woo, T.H.S., Smythe, L.D., Symonds, M., Norris, M., Dohnt, M.,
 Brenner, D.J. and Patel, B.K.C.
 Identification of Leptospiira inadai by continuously monitoring
 fluorescence during rapid cycle PCR
 TITLE
 JOURNAL
 REFERENCE
 2 (bases 1 to 288)
 Woo, T.H.S., Smythe, L.D., Symonds, M., Norris, M., Dohnt, M.,
 Brenner, D.J. and Patel, B.K.C.
 Direct Submission
 TITLE
 JOURNAL
 Submitted (24-MAR-1997) School of Science, Griffith University,
 Brisbane, QLD 4111, Australia
 FEATURES
 Location/Qualifiers
 source
 1..288
 /organism="Leptospiira inadai"
 /strain="218/95"
 /db_xref="taxon:29506"
 <1..>288
 /product="16S ribosomal RNA"
 77 a 61 c 88 g 62 t
 BASE COUNT
 ORIGIN

Query Match 85.5%; Score 18.8; DB 2; Length 288;
 Best Local Similarity 90.9%; Pred. No. 1.4e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 tgttgatcacagaatttgata 22
 | ||||| ||||| ||||| |||||
 Db 141 TATTGGATCACAGGATTGATA 162

RESULT 9
 LIU94979 288 bp DNA BCT 01-JAN-1998
 LOCUS
 DEFINITION
 Leptospiira inadai 16S ribosomal RNA gene, partial sequence.
 ACCESSION
 U94979
 VERSION
 U94979.1 GI:2735451
 KEYWORDS
 SOURCE
 Leptospiira inadai.
 ORGANISM
 Bacteria; Spirochaetales; Leptospiiraceae; Leptospiira.
 REFERENCE
 1 (bases 1 to 288)
 Woo, T.H.S., Smythe, L.D., Symonds, M., Norris, M., Dohnt, M.,
 Brenner, D.J. and Patel, B.K.C.
 Identification of Leptospiira inadai by continuously monitoring
 fluorescence during rapid cycle PCR
 TITLE
 JOURNAL
 REFERENCE
 2 (bases 1 to 288)
 Woo, T.H.S., Smythe, L.D., Symonds, M., Norris, M., Dohnt, M.,
 Brenner, D.J. and Patel, B.K.C.
 Direct Submission
 TITLE
 JOURNAL
 Submitted (24-MAR-1997) School of Science, Griffith University,
 Brisbane, QLD 4111, Australia
 FEATURES
 Location/Qualifiers
 source
 1..288
 /organism="Leptospiira inadai"
 /strain="1078 VRI"
 /db_xref="taxon:29506"
 <1..>288
 /product="16S ribosomal RNA"
 77 a 61 c 88 g 62 t
 BASE COUNT
 ORIGIN

Query Match 85.5%; Score 18.8; DB 2; Length 288;
 Best Local Similarity 90.9%; Pred. No. 1.4e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 tgttgatcacagaatttgata 22
 | ||||| ||||| ||||| |||||
 Db 141 TATTGGATCACAGGATTGATA 162

RESULT 10
 AB007012 353 bp DNA BCT 13-OCT-1997
 LOCUS
 DEFINITION
 Spirochaeta sp. 16S rRNA gene, partial sequence.
 ACCESSION
 AB007012
 VERSION
 AB007012.1 GI:2516253
 KEYWORDS
 16S ribosomal RNA.
 Spirochaeta sp. (sub-species: Freshwater obligate oligotroph,
 strain: SO-104) DNA.
 SOURCE
 Spirochaeta sp.
 ORGANISM
 Bacteria; Spirochaetales; Spirochaetaceae; Spirochaeta.
 REFERENCE
 1 (bases 1 to 353)
 Shin, M.-S.
 Direct Submission
 TITLE
 JOURNAL
 Submitted (05-SEP-1997) to the DDBJ/EMBL/GenBank databases. Mi-Sun
 Shin, Laboratory of Marine Molecular Microbiology, Faculty of
 Agriculture, Kyoto University; Oiwakecho, Kitashirakawa, Sakyo-ku,
 Kyoto, Kyoto 306-01, Japan (E-mail: misunekais.kyoto-u.ac.jp,
 Tel: 075-753-6224, Fax: 075-753-6226)
 2 (sites)
 Shin, M., Yoshinaga, I., Uchida, A. and Ishida, Y.
 Phylogenetic analysis by 16S rRNA gene sequencing of Obligate
 Oligotrophs isolated from the northern basin of Lake Biwa
 (Mesotrophic Lake)
 Unpublished (1997)
 JOURNAL

FEATURES	Location/Qualifiers	DEFINITION
source	1. .353	Arabidopsis thaliana chromosome II section 214 of 255 of the complete sequence. Sequence from clones T517.
	/organism="Spirochaeta sp."	AC003000 AB002093
	/strain="SO-104"	AC003000.2 GI:6598383
	/sub_species="Freshwater obligate oligotroph"	HTG
	/db_xref="taxon:28185"	thale cress.
rRNA	<1..>353	Arabidopsis thaliana
	/product="16S rRNA"	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
BASE COUNT	90 a 80 c 108 g 75 t	1 (bases 1 to 72590)
ORIGIN		Lin,X., Kaul,S., Rounsley,S.D., Shea,T.P., Benito,M.-I., Town,C.D., Fujii,C.Y., Mason,F.M., Bowman,C.L., Barnstead,M.E., Feldlyum,T.V., Buell,C., Ketchum,K.A., Lee,J.J., Ronning,C.M., Koo,H., Moffat,K.S., Cronin,L.A., Shen,M., Vanaken,S.E., Umayam,L., Tallon,L.J., Gill,J.E., Adams,M.D., Carrera,A.J., Creasy,T.H., Goodman,H.M., Somerville,C.R., Copenhaver,G.P., Preuss,D., Nierman,W.C., White,O., Eisen,J.A., Salzberg,S.L., Fraser,C.M. and Venter,J.C.
	Query Match 85.5%; Score 18.8; DB 1; Length 353;	Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana
	Best Local Similarity 90.9%; Pred. No. 1.4e+02;	Nature 402 (6763), 761-768 (1999)
	Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;	20083487
		10617197
QY	1 tgttgatcacaaagatttgata 22	2 (bases 1 to 72590)
		Lin, X.
Db	132 TGTGGATCACAGATCTGATA 153	Direct Submission
		Submitted (09-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA
RESULT 11		On Dec 17, 1999 this sequence version replaced gi:2642152.
ATHRNS2X	ATHRNS2X 1012 bp mRNA PLN 30-OCT-1994	The sequence and annotation of chromosome 2 were merged from those of the individual clones on this chromosome after removing overlaps. For detailed information, please see the TIGR web site (http://www.tigr.org/tdb/at/at.html).
LOCUS	Arabidopsis thaliana ribonuclease (RNS2) mRNA, complete cds.	Genes were identified by a combination of three methods: Gene prediction programs including GRAIL
DEFINITION		(ftp://arthur.epm.ornl.gov/pub/xgrail), Genefinder (Phil Green, University of Washington), Genscan (Chris Burge, http://gnomic.stanford.edu/GENSCANW.html), and NetPlantGene (http://www.cbs.dtu.dk/services/NetGene2/), searches of the complete sequence against a peptide database and plant EST databases at TIGR, and manual curations based on those analyses.
ACCESSION	M98336.1 GI:2829209	Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins
VERSION		Genes without protein or EST similarity, that are predicted by two or more gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats were identified by repeatmasker (Arian Smit, http://ftp.genome.washington.edu/RM/RepeatMasker.html). Genes are numbered from the top to bottom of the chromosome.
KEYWORDS	ribonuclease.	We thank the CSHL/WashU/ABI consortium for sequencing BAC clones F6P23, F5J6, T17A5, and T13L16, the ESSA group for sequencing clones F13D4, and Scott Jackson, Jiming Jiang, Klaus Meyer, Eric Richards and Satoshi Tabata for helpful assistance. In addition, we would like to thank the TIGR Bioinformatics Department, especially Lixin Zhou, Hanif Khalak, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy Peterson, Michael Holmes, and Delwood Richardson for software and database support.
SOURCE	Arabidopsis thaliana cDNA to mRNA.	This work was supported by the National Science Foundation, Department of Energy and the US Department of Agriculture.
ORGANISM	Arabidopsis thaliana	Address all correspondence to: at@tigr.org.
	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.	Location/Qualifiers
REFERENCE	1 (bases 1 to 1012)	1. .72590
AUTHORS	Taylor,C.B., Bariola,P.A., delCardayre,S.B., Raines,R.T. and Green,P.J.	/organism="Arabidopsis thaliana"
	RNS2: a senescence-associated RNase of Arabidopsis that diverged from the S-RNases before speciation	/db_xref="taxon:3702"
TITLE		16. .795
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 90 (11), 5118-5122 (1993)	/gene="RNS2"
MEDLINE	93281708	16. .795
FEATURES	Location/Qualifiers	/gene="RNS2"
source	1. .1012	/codon_start=1
	/organism="Arabidopsis thaliana"	/product="ribonuclease"
gene	16. .795	/protein_id="AAA51406.1"
CDS	16. .795	/db_xref="GI:289210"
	/translation="MASRLGILLVACIAGAFAGDVIELNRSQREDFYFALSLOWPCT"	YCRTRHCSKNACRCGSDAPOTFIHGLWFDYNDGSPCCYRSDFEKEIETLMDG
	YCRTRHCSKNACRCGSDAPOTFIHGLWFDYNDGSPCCYRSDFEKEIETLMDG	LEKYWPSICGSPSSCGSGKSGFHEWKEHGTCSVPFHDEYNVFLTLNLKHN
	TDVLYQAGYVANSSEKYPGLGIVTAIONAFHITPEVYCKRDADIEIRICFYKDFKPRD	CVSODLTSRKSCPKYVSLPEYTPLDGEAWLKMPTREAL"
BASE COUNT	287 a 197 c 217 g 311 t	
ORIGIN		
	Query Match 85.5%; Score 18.8; DB 7; Length 1012;	
	Best Local Similarity 90.9%; Pred. No. 1.3e+02;	
	Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;	
QY	1 tgttgatcacaaagatttgata 22	
Db	672 TGTGGTTCACAGATTGACA 693	
RESULT 12		
AC003000	AC003000 72590 bp DNA PLN 05-APR-2000	
LOCUS		

Arabidopsis thaliana chromosome II section 214 of 255 of the complete sequence. Sequence from clones T517.

AC003000 AE002093
AC003000.2 GI:6598383

HTG.
thale cress.

Arabidopsis thaliana

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

1 (bases 1 to 72590)

Lin,X., Kaul,S., Rounsley,S.D., Shea,T.P., Benito,M.-I., Town,C.D., Fujii,C.Y., Mason,T.M., Bowman,C.L., Barnstead,M.E., Feldblum,T.V., Buell,C.R., Ketchum,K.A., Lee,J.J., Ronning,C.M., Koo,H., Moffat,K.S., Cronin,L.A., Shen,M., VanAken,S.E., Unayam,L., Tallon,L.J., Gill,J.E., Adams,M.D., Carrera,A.J., Creasy,T.H., Goodman,H.M., Somerville,C.R., Copenhaver,G.P., Preuss,D., Nierman,W.C., White,O., Eisen,J.A., Salzberg,S.L., Fraser,C.M. and Venter,J.C.

Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana

Nature 402 (6763), 761-768 (1999)

20083487

10617197

2 (bases 1 to 72590)

Lin,X.

Direct Submission

Submitted (09-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA

On Dec 17, 1999 this sequence version replaced gi:2642152.

The sequence and annotation of chromosome 2 were merged from those of the individual clones on this chromosome after removing overlaps. For detailed information, please see the TIGR web site (<http://www.tigr.org/tdb/at/at.html>).

Genes were identified by a combination of three methods: Gene prediction programs including GRAIL

(<ftp://arthur.epm.ornl.gov/pub/xgrail>), Genefinder (Phil Green, University of Washington), Genscan (Chris Burge,

<http://gnomic.stanford.edu/GENSCANW.html>), and NetPlantGene (<http://www.cbs.dtu.dk/services/NetGene2/>), searches of the

complete sequence against a peptide database and plant EST databases at TIGR, and manual curations based on those analyses.

Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins.

Genes without protein or EST similarity, that are predicted by two or more gene prediction programs over most of their length are

annotated as 'hypothetical' proteins. Genes encoding tRNAs are

predicted by tRNAscan-SE (Sean Eddy,

<http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats were

identified by repeatmasker (Arian Smit,

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>). Genes are

numbered from the top to bottom of the chromosome.

We thank the CSHL/WashU/ABI consortium for sequencing BAC clones F6P23, F5J6, T17A5, and T13L16, the ESSA group for sequencing clone F13D4, and Scott Jackson, Jiming Jiang, Klaus Meyer, Eric Richards

and Satoshi Tabata for helpful assistance. In addition, we would like to thank the TIGR Bioinformatics Department, especially Lixin Zhou, Hanif Khailak, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy Peterson, Michael Holmes, and Delwood Richardson for software and

database support.

This work was supported by the National Science Foundation, Department of Energy and the US Department of Agriculture.

Address all correspondence to: at@tigr.org.

Location/Qualifiers

1. .72590

/organism="Arabidopsis thaliana"

/cultivar="Columbia"

FEATURES
source

/db_xref="taxon:3702"
/chromosome="II"
1. >72590
/note="Sequence from clone T517"
complement(130.198)
/rpt_family="(CAAA)n"
complement(1000.1024)
/rpt_family="POLY_A"
complement(1266.1291)
/rpt_family="(CAAAA)n"
complement(2759.2862)
/rpt_family="POLY_A"
complement(4584.5909)
/gene="At2g39710"
complement(4584.5909)
/gene="At2g39710"
complement(4584.5909)
/note="T517.1; similar to
GI12245012|gql|PID|e327000|297341"
complement(4584.5909)
/gene="At2g39710"
/note="unknown protein"
/codon_start=1
/protein_id="AAB87120.1"
/db_xref="GI:2642153"
/translation="MDSSSSSSLSKFLRISVLLIPLPFLFKTSSTNOTLPLSLK
TOKLPOSSDKLSFRNVLTLVAVGDPQNIWLDVGLSLSLHCKKSPNLSGVF
NPVSSSYSPVPCSSP|ICRTRRDLIPASCDPKTHLCHVAISYADATSEGNLAHET
FVIGSVTRGTLFGCNDSSGSLNSSEDAKSTGLMGNRCSLSFVNLQGFSKSYLSLPK
SDSSGFLLLGSDYSWLGPIQYPLVQSTPLPYFDRVAYTVQLEGIRGVSKYLSLPK
SVFVDPHTGAGOTVDSGQFTFLMGVPVYTLKNEFTQTVSLRVLDDPDPVFGTGM
DLCKYGVSTRNPSGLPMVSLMFRGAEMSVGOKLLYRVNGAGSGEGKEEVCYFTFGN
SDLLGTEAFVIGHHQQNMWMEFDLAKSRVGFAGNVRCDLASQRLGRD"
complement(<10039.11304)
/gene="At2g39720"
complement(<10039.11304)
/note="T517.2"
/gene="At2g39720"
complement(10039.11244)
/gene="At2g39720"
/note="unknown protein"
/codon_start=1
/protein_id="AAB87121.1"
/db_xref="GI:2642154"
/translation="MASGYSWCYSCSRFVWVSDSISCPDCDGGFLELQEPDLFTPSD
SFTTTTQHRSTREPPSSSTPSASHADNSPTTIVTRNSRSPNVIVILRG
SAAATPSDVVSEGLDRFAQMYDDGTGSLRPLPSPMEFLIGSGFDRLLDQISOIE
LNTNLRNRSCEHPSPAKSAIEALPLIEIDPTLLSDSQSHCAVCNFKVLSKAREMP
CNHLYHPDCILPLAIRNSCPVCRHELPAEDLDTGTGAALTAVTAAEEEDSAAAGLT
IWRLPGGFAVGRIPGSGWRGDRMPVYTVEDGRLGDERLPRVAVGWSRRGRDGG
GSRERGGFAGRIHRLFGCFSGSGSGSIAAAAASSGSGSRIRVTRTRFSMFSTASS
SSRRRNWLA"
complement(11453.11512)
/rpt_family="POLY_A"
complement(join(13323.13599,13724.14005,14087.14559,
14651.14736,14836.14941,15033.15349,15834.16055))
/gene="At2g39730"
complement(join(13323.13599,13713.14005,14087.14559,
14651.14736,14836.14941,15033.15349,15834.16055))
/gene="At2g39730"
complement(13323.16055)
/gene="At2g39730"
/note="T517.3"
13364.13461
/rpt_family="POLY_A"
complement(join(13484.13599,13724.14005,14087.14559,
14651.14736,14836.14941,15033.15349,15834.15878))
/gene="At2g39730"
/note="alternative splice form 1 GP1166834|M86720"
/codon_start=1
/product="Rubisco activase"
/protein_id="AAB87122.1"
/db_xref="GI:2642155"
/translation="MAAAVSTVGAINRAPLSLNGSGGSAVASPASTFLGKKVTVSRF

AQSNKNSGSKFLAVKEDKQTDGDRWRGLAYDTSDDQDITRKGMDVDSVFOAPMGT
GTHHVLSSYVYVSOGLRVNLDNMMDGFYIAPAFMDKLVVHITKFLTNPKVPLI
LGIWGGKQCKSQFQELVMAKMGINPTMSAGELESGNAGEPAKLIQRORYREAADLIK
KGMKCLFINDLDAGAGMGCTTQYVNNQMVNATLMNIADNPVQLPCMYKNEENA
RVPLICTGNDFSTLYAPLIRDRMEKFEYWAPTREDRIGVCKGIERTDKIKDEDIVTLV
DQFPGQSIDFFGALRARVYDDEVKFEVSGEIKRRLVNSREGPPVFEQPEPTYEK
LMEYGNLMVMEQENVRVQLAETYLSQLAALGDANADAIGRTFYGKGAAQOVNLPVPEG
CTDPVAENFDPTARSDDGTCVNF"
complement(join(13579.13599,13713.14005,14087.14559,
14651.14736,14836.14941,15033.15349,15834.15878))
/gene="At2g39730"
/note="alternative splice form 2 GP1166835|M86720"
/codon_start=1
/product="Rubisco activase"
/protein_id="AAB871137.1"
/db_xref="GI:2642170"
/translation="MAAAVSTVGAINRAPLSLNGSGGSAVASPASTFLGKKVTVSRF
AQSNKNSGSKFLAVKEDKQTDGDRWRGLAYDTSDDQDITRKGMDVDSVFOAPMGT
GTHHVLSSYVYVSOGLRVNLDNMMDGFYIAPAFMDKLVVHITKFLTNPKVPLI
LGIWGGKQCKSQFQELVMAKMGINPTMSAGELESGNAGEPAKLIQRORYREAADLIK
KGMKCLFINDLDAGAGMGCTTQYVNNQMVNATLMNIADNPVQLPCMYKNEENA
RVPLICTGNDFSTLYAPLIRDRMEKFEYWAPTREDRIGVCKGIERTDKIKDEDIVTLV
DQFPGQSIDFFGALRARVYDDEVKFEVSGEIKRRLVNSREGPPVFEQPEPTYEK
LMEYGNLMVMEQENVRVQLAETYLSQLAALGDANADAIGRTFYGKGTEEPEPSK"
complement(17561.17639)
/rpt_family="(TA)n"
join(<18783.18921,19188.19352,19496.19674,19766.19861,
19965.20034,20131.20261,20593.>21237)
/gene="At2g39740"
complement(18783.21237)
/note="T517.4; predicted by genscan"
join(18783.18921,19188.19352,19496.19674,19766.19861,
19965.20034,20131.20261,20593.>21237)
/gene="At2g39740"
/note="hypothetical protein"
/codon_start=1
/protein_id="AAB87123.1"
/db_xref="GI:2642156"
/translation="MSRNPFLDPTLQELIOLVQIKTRADRTITVIDOLRDVLOSVEC
LRGATVQPGSFVSNLFTRMGDLDSVDFSGSSILFTGKKQKOTLLGHLRLRASP
LWYKLOFVIHARVPLIKVYVSGHQIRSCDISIDNLDGLKSRFLFWISIDGRDVLV
LVKEWAKHININDSKTGFNSYLSLVIFHFQTCVPAIILPLRVLVYPSKASRDLTVG
RTADESTAQVTAANIRKSRKSVNRSSSELLVSEFFAKVDFPQPVNAARSV
RNLDRIAQVFOITSRRLVSECNRSIIIGLTQHIQIESLYRTISLPSQHANGHNH
RNLHGAEPQNOQMOQNSOSYNTNPHPWPLTQSRPQONNPNRNLOQPPVQ
QTWVITQTOQCKSPYKSGNPLKNTSAGSQNGHIGKPKSGHNGVNSARPATNG
VNSARPPSPKIPSGGGIWRPRHEQ"
complement(join(21380.21726,21839.21916,22124.22246,
22328.22585,22834.22993,23212.23322,23481.23751,
23847.23996,24089.>24814))
/gene="At2g39750"
complement(21380.>24814)
Query Match 85.5%; Score 18.8; DB 6; Length 72590;
Best Local Similarity 90.9%; Pred. No. 85;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 tgttggtacacagatttgata 22
|||||-----|||
Db 35807 TGTGGTTCACAGATTTGACA 35828
RESULT 13
AF165146 109723 bp DNA HTG 01-JUN-2000
LOCUS Homo sapiens chromosome 8 map 8q12-8q13 clone CTA-397H3, WORKING
DEFINITION DRAFT SEQUENCE, 3 unordered pieces.
ACCESSION AF165146
VERSION AF165146.2 GI:8151945
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

1 (bases 1 to 109723)

REFERENCE
AUTHORS Schudy,A., Blechschmidt,K., Menzel,U., Polley,A., Reichwald,K., Rump,A., Schillabel,M.B., Taudien,S., Wen,G., Schlegelberger,B., Siebert,R. and Rosenthal,A.

TITLE Chromosome 8 genomic sequence

JOURNAL Unpublished

REFERENCE
AUTHORS Schudy,A., Blechschmidt,K., Schillabel,M., Baumgart,C., Menzel,U., Weber,J., Schattevoy,R. and Rosenthal,A.

TITLE Direct Submission

JOURNAL Submitted (06-JUL-1999) Genome Analysis, Institute of Molecular Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany

REFERENCE
AUTHORS Schudy,A., Blechschmidt,K., Schillabel,M., Baumgart,C., Menzel,U., Weber,J., Schattevoy,R. and Rosenthal,A.

TITLE Direct Submission

JOURNAL Submitted (24-MAY-2000) Genome Analysis, Institute of Molecular Biotechnology, Beutenbergstr. 11, Jena 07745, Germany

COMMENT

On Jun 1, 2000 this sequence version replaced gi:5514637.

* NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

* 1 61961: contig of 61961 bp in length

* 61962 62061: gap of unknown length

* 91746: contig of 29685 bp in length

* 91747 91846: gap of unknown length

* 91847 109723: contig of 17877 bp in length.

Location/Qualifiers

1. 109723

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="8 map 8q12-8q13"

/clone="CTA-397H3"

misc_feature

1. 61961

/note="assembly_fragment"

clone_end:SP6

vector_side:left"

misc_feature

91647..109723

/note="assembly_fragment"

clone_end:T7

vector_side:right"

BASE COUNT 31869 a 22757 c 23188 g 31709 t 200 others

ORIGIN

Query Match 85.5%; Score 18.8; DB 72; Length 109723;

Best Local Similarity 90.9%; Pred. No. 82;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 tgttggtacacagatttgata 22

1 |||||

DB 95330 TTTTGGATCACAAAGATTGGTA 95351

RESULT 14

AF137379 AF137379 200799 bp DNA circular PLN 14-SEP-1999

LOCUS Nephroselmis olivacea chloroplast DNA, complete genome.

DEFINITION Nephroselmis olivacea chloroplast DNA, complete genome.

ACCESSION AF137379

VERSION AF137379.1 GI:5880684

KEYWORDS

SOURCE Nephroselmis olivacea.

ORGANISM Eukaryota; Viridiplantae; Chlorophyta; Prasinophyceae; Chlorodendraceae; Chlorodendraceae; Nephroselmis.

REFERENCE 1 (bases 1 to 200799)

AUTHORS Turmel,M., Otis,C. and Lemieux,C.

TITLE The complete chloroplast DNA sequence of the green alga Nephroselmis olivacea: insights into the architecture of ancestral chloroplast genomes

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (18), 10248-10253 (1999)

MEDLINE 99398694

REFERENCE 2 (bases 1 to 200799)

AUTHORS Turmel,M., Otis,C. and Lemieux,C.

TITLE Direct Submission

JOURNAL Submitted (24-MAR-1999) Biochimie, Universite Laval, Pavillon Charles-Eugene Marchand, Quebec G1K 7P4, Canada

FEATURES

Location/Qualifiers

1. 200799

/organism="Nephroselmis olivacea"

/organelle="plastid:chloroplast"

/db_xref="taxon:31312"

/cell_line="NIES 484"

misc_feature

1. 92126

/note="Region: large single-copy region (LSC)"

55..137

/gene="trnL(gag)"

/note="codons recognized: CUY"

/product="tRNA-Leu"

/anticodon="(pos:88..90,aa:Leu)

55..137

/gene="trnL(gag)"

332..634

/gene="rps14"

332..634

/gene="rps14"

/codon_start=1

/product="ribosomal protein S14"

/protein_id="AAD54778.1"

/db_xref="GI:5880685"

/translation="MAKSMIERESKRAALVAKYATRRQALKAAIQKTSFDERLVLQ

HQLQDLPVNSVPCRLHNRCTITGRPKGYRFGLSRHELRAMHGCLLPGVTKASW"

765..849

/gene="trnL(uaa)"

/note="codon recognized: UUA"

/product="tRNA-Leu"

/anticodon="(pos:798..800,aa:Leu)

765..849

/gene="trnL(uaa)"

983..1054

/gene="trnQ(uug)"

/note="codons recognized: CAR"

/product="tRNA-Gln"

/anticodon="(pos:1015..1017,aa:Gln)

983..1054

/gene="trnQ(uug)"

1102..1182

/gene="trnL(uag)"

/note="codons recognized: CUR"

/product="tRNA-Leu"

/anticodon="(pos:1136..1138,aa:Leu)

1102..1182

complement(1364..1468)

/gene="psbM"

complement(1364..1468)

/gene="psbM"

/codon_start=1

/product="M protein of photosystem II"

/protein_id="AAD54779.1"

/db_xref="GI:5880686"

/translation="MEVNILGIATLFIIPTSFLILLYVKVTASQON"

1618..1691

/gene="trnM(cau)"

/note="codon recognized: AUG"

/product="tRNA-Met"

/anticodon="(pos:1652..1654,aa:Met)

1618..1691

/gene="trnM(cau)"

1733..3862

```

CDS
/gene="ftsI"
/locus="1733..3862"
/gene="ftsI"
/codon_start=1
/product="putative peptidoglycan synthetase"
/protein_id="AAD54780.1"
/db_xref="GI:5880687"
/translation="M1YNYPMKYRROFRRLPKHLOGIYVAFISISMIKIALDPYSKR
PKMWSGTPROVENRMWIKLSLATVGLFATRLSGLOFNKYTELKSAVERQOIGQ
VNDPQQRKILDHGDIVAIDLPAYDLYVHPKMSLSERIAELLSLDLSQYLYN
RLDEGSGICHOIDTNTSAQIRRLGVGDIELVHPQRYPKRGSFESILGTVDEG
YGOAGLESSDDMMKSTYQDPCWMDGHGFLIRFPKQILFHQESALQITDCGLQE
KVSQLTNNAMRRGAKRIAI IEAHSGAIRCLATSPYDPCNYGFWPFMERCPWIT
DLPEGSTFKPNALALENGIFQPTDRILDITKIRIGDSWIGNVGGFTWDRSLDHL
TGQIILQRSNVMVRMOSLDPAIYHRLNIRLGLGSHRNDNOTSEKMSHSDHNEGSM
NLADLSDVALSVDQDEFVDEHEIEAATASGQGLAMTPLKLIQIATANGMAVT
PHLSKIVTLDFHHQLQSNNEFSLQGVGSLRSQYHAKQRPRTYTHDLYLGHVPVP
SLELGFVDKSPHPTREKRRLFSROTNCNVLQEQVLDQATGSRGLFPGFANAG
KTGAQKASALGGYSTSVTSFVGIYPVAKPKFVTLVIIDEPDPFRFGNTAVDVT
QTLISEMIVQEQPPSYPTVSLFRNN"
3998..5059
/gene="psbA"
/codon_start=1
/product="D1 reaction-center protein of photosystem II"
/protein_id="AAD54781.1"
/db_xref="GI:5880688"
/translation="MTAILRRRESTSWARFCDMVTSTENRLYIGWFGVLMIPLLLTA
TSVFIIGFIAAPVDIDGIEPVPVSGSLICNNILISGAI LPSSAAIGIHVPIWEAASI
DEWLINGGCVELIVLHFLGACVIMGRENWELSPRLCHRPWIAVAYAPVAATAVAVLI
YPIQGSFSDMGIGISGTFNFMVFOAEHNLIMHPHMLGVAGVFGSGLSFSAMHGSLL
VTSLSIRTFENESANAGYFGQEEETYNIAAAGYFGRGLFOYAFSNRSRSLHFFLA
AMPVVCITFALGVSMFNLGNGFNFNQSVVDQSGRVINTWADIINRANLGMVEVMHER
NANFPLDLASVDAPAVQG"
5489..6964
/gene="atpB"
/codon_start=1
/product="CF1 beta subunit of ATP synthase"
/protein_id="AAD54782.1"
/db_xref="GI:5880689"
/translation="MSQAGVKTKQAGKIVQIVGPMVDAVAPQSGMPNINIALIVKGRN
GAGEVSVQCEVQOGLDGLVRAVMSATDGLMRGMEVTDTRALSVPVGPPTLGRIF
NVLGEVDNMGVGNKTEPLIHRAPAVDLDTKLSIFETGIKVDVLLAPYARRGGKIG
LFGAGVGKTVLMLNNIAKAGHGVGFGVGEREGNDLYMEMCESKYRNKADY
KSSKVALVYQGMNEPPCARMGVLTALTMAEYFRDVNNQDVLFDINIFRFVQAGSEV
SALLGRMPSAVGQPTLATETMGLOERITSTKQGSITSIQAVVVPADDLTDPAATTF
AHLDATVLSRNLAAGIYPADVPDLSTSTMQLPWLIGDDHYGTAQVKOTLQRYKEL
ODIITAILGDELSEEDRMIVARAKKIERFLUSQPFVFAEVFTGAPKIVPLAESIQGFK
LILSGELSLPESAFYLVGNIEAIAKAAASIQAAAAA"
6977..7378
/gene="atpE"
/codon_start=1
/product="CF1 epsilon subunit of ATP synthase"
/protein_id="AAD54783.1"
/db_xref="GI:5880690"
/translation="MSLQVRLTPEQVFLNVADEIILPTNTGOMGVLTNHTPLITAI
DIGPMVRSSTWQSMALLGGLALVKDNOVILVNEAELGSDINAEAEETFLAAKKA
LANSKTRDQENNLAFKRVRYQATLVK"
complement(7622..8104)
/gene="petB"
complement(7622..8104)
/gene="petD"
/codon_start=1
/product="subunit IV of cytochrome b6/f complex"
/protein_id="AAD54784.1"
/db_xref="GI:5880691"
/translation="MSVTKKPDLTDPVLRAKLAKGMGHNYGEPAPNDLLIYMPVVI
LGTLSCTGLAVLDPAAIGEPANPEATPLEILPEWYFFPVQLLRTVPNKLGLVLLMA
AVPAGLLTVPFIESINKFQNPFRPVATTVFILGTVVVAIWLIGATLPIDISLTGLGF"
gene
complement(8132..8779)
/gene="petB"
CDS
complement(8132..8779)
/gene="petB"
/codon_start=1
/product="apocytochrome b6 of cytochrome b6/f complex"
/protein_id="AAD54785.1"
/db_xref="GI:5880692"
/translation="MSKIYDWEERLEIQAIADDTSKYVPPHVNIFYCFGGTILTCF
LIQVATGFAMTYRTVTVEAFASVEYIMTNVNFGLIRSIHRWSASMMVMLILHVF
RYLVTGGFKPRELTVTAVILAVITVSFGVTGYSLPDQVGYWAVKIYTVGPDALPV
IGAPLVELLGSVSGQSLTRFYSLSLHTEVLPVLLTAVFMLHFLMIRKOGISGPL"
complement(9045..9431)
/gene="rpl12"
complement(9045..9431)
/gene="rpl12"
/codon_start=1
/product="ribosomal protein L12"
/protein_id="AAD54786.1"
/db_xref="GI:5880693"
/translation="MSTITNQIEOLKSMTLLEAAELVSOIETFCVDASAPAAGAVM
VAAAPSPVVEEQTEFTVINDVPSAKRIAVIKVVRALTGLCLKEAKDMIESVPKAI
KEGICKDEAQOIKQOLEEAGATVTK"
complement(9442..9846)
/gene="rps9"
complement(9442..9846)
/gene="rps9"
/codon_start=1
/product="ribosomal protein S9"
/protein_id="AAD54787.1"
/db_xref="GI:5880694"
/translation="MDHTSSVGLRKTGRKCAVAQVRIIEGTQQLTINQRPVGYLOY
NPAYLLAAGALDVTETLNTYDTIVTKGGTLTGQAEIKLGLARALCSLIHKRKAL
KPGYLTRNPLRKRKKYGLKKAKAPQFSKR"
complement(9998..11485)
/gene="rpoA"
complement(9998..11485)
/gene="rpoA"
/codon_start=1
/product="alpha subunit of RNA polymerase"
/protein_id="AAD54788.1"
Query Match 85.58; Score 18.8; DB 6; Length 200799;
Best Local Similarity 90.94; Pred. No. 77;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 ttttgatcacacaagtattgata 22
||||| ||||| ||||| |||||
Db 134224 TGTGGCTCACATGATTGATA 134245
RESULT 15
AF137379/c
LOCUS AF137379 200799 bp DNA circular PLN 14-SEP-1999
DEFINITION Nephroselmis olivacea chloroplast DNA, complete genome.
ACCESSION AF137379
VERSION AF137379.1 GI:5880684
KEYWORDS Nephroselmis olivacea.
SOURCE Nephroselmis olivacea.
ORGANISM Chloroplast Nephroselmis olivacea
Eukaryota: Viridiplantae; Chlorophyta; Prasinophyceae;
Chlorodendres; Chlorodendraceae; Nephroselmis.
REFERENCE 1 (bases 1 to 200799)
AUTHORS Turmel,M., Otis,C. and Lemieux,C.
TITLE The complete chloroplast DNA sequence of the green alga
Nephroselmis olivacea: insights into the architecture of ancestral
chloroplast genomes
Proc. Natl. Acad. Sci. U.S.A. 96 (18), 10248-10253 (1999)
MEDLINE 99398694
REFERENCE 2 (bases 1 to 200799)
AUTHORS Turmel,M., Otis,C. and Lemieux,C.

```

TITLE Direct Submission
JOURNAL Submitted (24-MAR-1999) Biochimie, Universite Laval, Pavillon
Charles-Eugene Marchand, Quebec G1K 7P4, Canada
FEATURES Location/Qualifiers
source 1. 200799
/organism="Nectroselmis olivacea"
/organelle="plastid:chloroplast"
/db_xref="taxon:31312"
/cell_line="NIES 484"
misc_feature 1. 92126
/note="Region: large single-copy region (LSC)"
tRNA 55..137
/gene="trnL(gag)"
/note="codons recognized: CUY"
/product="tRNA-Leu"
/anticodon=(pos:88..90,aa:Leu)
gene 55..137
/gene="trnL(gag)"
332..634
/gene="rps14"
332..634
/gene="rps14"
/codon_start=1
/product="ribosomal protein S14"
/protein_id="AAD54778.1"
/db_xref="GI:5880685"
/translation="MARKSMIERESKRAALVAKYATRRQALKAAIQKTSFDERLVIQ
HQQLDVPNSVPCRLHNRCTTGRPKGYDFGLSRHRLAMAGCULLPGVTRASW"
tRNA 765..849
/gene="trnL(uaa)"
/note="codon recognized: UUA"
/product="tRNA-Leu"
/anticodon=(pos:798..800,aa:Leu)
gene 765..849
/gene="trnL(uaa)"
983..1054
/gene="trnQ(uug)"
/note="codons recognized: CAR"
/product="tRNA-Gln"
/anticodon=(pos:1015..1017,aa:Gln)
gene 983..1054
/gene="trnQ(uug)"
1102..1182
/gene="trnL(uag)"
/note="codons recognized: CUR"
/product="tRNA-Leu"
/anticodon=(pos:1136..1138,aa:Leu)
1102..1182
/gene="trnL(uag)"
complement(1364..1468)
/gene="psbM"
complement(1364..1468)
/gene="psbM"
/codon_start=1
/product="M protein of photosystem II"
/protein_id="AAD54779.1"
/db_xref="GI:5880686"
/translation="MEYNILGLIATLFIITPSFLLILYVKTASQON"
1618..1691
/gene="trnM(cau)"
/note="codon recognized: AUG"
/product="tRNA-Met"
/anticodon=(pos:1652..1654,aa:Met)
1618..1691
/gene="trnM(cau)"
1733..3862
/gene="ftsI"
1733..3862
/gene="ftsI"
/codon_start=1
/product="putative peptidoglycan synthetase"
/protein_id="AAD54780.1"
/db_xref="GI:5880687"

/translation="MIYNPMKYRRQRRLPKHLQGIYYAFLSTSTWIKIALDPSYKR
PMKMHSGTFFQYENRMVMIKSLATVGLLFATRLSGLOFNKYTELKSYVAERQOIG
VNDQQRKIIILDHGDIVADLPAYDLYVHPRMCSISLERIAELLSPLIDLSOQLYN
RLDEGDSICLMOHIDNTSAQIRRLGVDGIELVHPQVYVPRKGSFSLGYVDTEG
YGOAGLESSLDDNMKSTYQDVPCMHGDNFLGIRFPKQILFHQESALQLTIDCGLQE
KVSOLITNANRFGAKRIAIIMEAHSGAIRCLATSPSYDNCYGVWPFMRFCWPIT
DLFEPGSGTFKPNLAIVALENGIFQPTDRILDTGIRIGDSWIGNVGGFIDWRLDHL
TGTOILQORSNVGMVYRQSLDPAIYHRLIRLGLGRNDNQTSFKMSHDHNSGW
NLKDLTSDYATISVVKDQDEFVDHEIEAATASFGOGLAMTPLKLLQLLATLANGMAVT
PHLISKIVTLDFHHLHMFESLQMGVOSVLSRSOVHAKQPRPYTHDLYLGHVPVP
SLELGWFDVKSIPHTRRERLFSROTCNVLLGNLEOVLDAAOTSGRGLPGVAMAG
KTGTAKASALGCGYSTDSVTYFVGIYPANKFVTLVIIDEDPDPFRFGENTAVDT
QTLLSEMIIVQEQDPPSTPTVSLPERNM"
3998..5059
/gene="psbA"
3998..5059
/gene="psbA"
/codon_start=1
/product="D1 reaction-center protein of photosystem II"
/protein_id="AAD54781.1"
/db_xref="GI:5880688"
/translation="MTAILERRESTSVARFCDMVTSTENRLYIGWFGVLMIPLLLLTA
TSVFLIGFIAAPPVDIDGIREPVSGLLFGNNIISGAIIPSSAAIGIFHYPIWEAASI
DEWLYNGCYELIVLHLLGVACVMGREWELSPRLGMRPTAVAYSAPAAATAVFLI
YPIGSGFSOGMPLGISGTENFMIVQAEHNLHMPFHMUGVAGVFGSLSFSAHGSLL
VTSLIETTESANAGYKFGQBEETYNIVAAHGIFGRLIFQYAFSNNSSLHFLLA
AMPVVCITWFTALGVSTWAFNLGNFNQSVVDQGRVINTWADIINRANLGMVWHER
NAHNFPLDASVDAPAVQG"
5489..6964
/gene="atpB"
5489..6964
/gene="atpB"
/codon_start=1
/product="CF1 beta subunit of ATP synthase"
/protein_id="AAD54782.1"
/db_xref="GI:5880689"
/translation="MSQAGVYTKQAGKIVQIVGPMVDVAFQPSMPNIYNLIYKGRN
GAGQEVSVCEVQQLGLDGLVRAVMSATDGLMRGMEVTDTRALSVPVGTTLGRIF
NVLGEVDNMGVGNKTLPIHREAPAFVLDLTKLISFETGIKVYDILLAPVRRGKIG
KFGAGVGKTVLIMELINNTAKAGHVSFGVGGERNDGLYMENCESVKVINKADV
LSSKVALVQGMNEPPGARMRVGLTALTMAEYFRDVANNQDVLFFIDNIFRVOAGSEV
SALLGRMPSAVGYOPTLATENGLOERTSTKDGSIISIOAVYVPADDLTDPAATTF
AHLDTTVLSRNLAAKGIYPAYVDPLDSTMLQPMILGDDHYGTAQRVKQLRYKEL
QDIIAIGLDLSEEDMIVARARKIERLSFPFVAEFTGAPGKYVPLAESIQGKEL
LILSGELDSLPESAFYLVGNITEAIAKASIAQAAAAA"
6977..7378
/gene="atpE"
6977..7378
/gene="atpE"
/codon_start=1
/product="CF1 epsilon subunit of ATP synthase"
/protein_id="AAD54783.1"
/db_xref="GI:5880690"
/translation="MSLQVRLLTPQVFLNVSADLILPTNTGOMGVLTNHTPLITAI
DIGPMLYRSESTWQSMALLGLALVKNQVILVNEAELGSDINAEAEATTFLLAKEA
LANSKTRKQIENNLAFKRAKRYQVATLVK"
complement(7622..8104)
/gene="petD"
complement(7622..8104)
/gene="petD"
/codon_start=1
/product="subunit IV of cytochrome b6/f complex"
/protein_id="AAD54784.1"
/db_xref="GI:5880691"
/translation="MSVTKKPDLTDPVLRAKLAKMGHNYGYGEPAWPNLLLYMFPVVI
LGTLSCTIGLAVLDPAAGPAPNPATPLEILPEWYFPVQLLRTVPNKLGLVLLMA
AVPAGLLTVPFIESINKFQNPFRPVATTVFLIGIVVAIMLGIGATLFDISLTGLFL
"
complement(8132..8779)
/gene="petB"
complement(8132..8779)
/gene="petB"
/codon_start=1

```

/product="apocytochrome b6 of cytochrome b6/f complex"
/protein_id="AAD54785.1"
/db_xref="GI:5880692"
/translation="MSKIYDFEERLEIQAIADDTTSKYVPPHVNIFYCFGGITLTCTF
LIQVATGFAMTYFRVTTEAFASVEYIMTNVFWGLIRSIHRWSASMMVMLLIHVF
RVLTGGRFKPRELWTVGVILAVITVSEFVGYSLPQDQGYWAVKIIVTGPDAIPV
IGAPLVLLRGSVSGQSTLTFRFYSLHTEVLPLLTAVFMLHFLMIRKQGISGPL"
complement(9045..9431)
/gene="rpl12"
complement(9045..9431)
/gene="rpl12"
/codon_start=1
/product="ribosomal protein L12"
/protein_id="AAD54786.1"
/db_xref="GI:5880693"
/translation="MSTITNQIIIEOLKSMTLLEAAELVSOIETFGVDASAPAAGAVM
VAAAPSAFVVEQTEFTVMINDVPSAKRIAVIKVVRALTGLGLREAKDMIESVPKAI
KEGIGKDEAQQINQQLLEEAGATVTR"
complement(9442..9846)
/gene="rps9"
complement(9442..9846)
/gene="rps9"
/codon_start=1
/product="ribosomal protein S9"
/protein_id="AAD54787.1"
/db_xref="GI:5880694"
/translation="MDHTSSVGLRKTGRKCAVAOVRIIEGTQLTINORPGVPYIQY
NPAYLLAOCALDVTELTNRYDTIVKTGGGLTGQAEAIKGLARALCSLIHKHKAL
KPGYLTRNPLRKRKKYGLKKARKAPQPSKR"
complement(9998..11485)
/gene="rpoA"
complement(9998..11485)
/gene="rpoA"
/codon_start=1
/product="alpha subunit of RNA polymerase"
/protein_id="AAD54788.1"

```

```

Query Match      85.5%  Score 18.8; DB 6; Length 200799;
Best Local Similarity 90.9%  Pred. No. 77;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

Qy 1 tgttgatcacagaatttgata 22
      ||||| ||||| ||||| |||||
Db 158702 TGTGGCTCACATGATTGATA 158691

```

```

Search completed: March 27, 2001, 08:17:03
Job time: 5900 sec

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 27, 2001, 06:38:43 ; Search time 2286.42 Seconds
(without alignments)
3306.005 Million cell updates/sec

Title: US-09-380-826A-1
Perfect score: 1477
Sequence: 1 gatcatggctcagaactaac.....ccgtaaatcgattcctgcag 1477

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1118133 seqs, 2558875100 residues 2236266
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba1.*
- 2: gb_ba2.*
- 3: gb_om.*
- 4: gb_ov.*
- 5: gb_ph.*
- 6: gb_pl1.*
- 7: gb_pl2.*
- 8: gb_pr1.*
- 9: gb_pr2.*
- 10: gb_pr3.*
- 11: gb_ro.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: em_fun.*
- 15: em_hum1.*
- 16: em_hum2.*
- 17: em_in.*
- 18: em_om.*
- 19: em_or.*
- 20: em_ov.*
- 21: em_pat.*
- 22: em_ph.*
- 23: em_pl.*
- 24: em_ro.*
- 25: em_sts.*
- 26: em_sy.*
- 27: em_un.*
- 28: em_v1.*
- 29: gb_ba3.*
- 30: gb_in1.*
- 31: gb_in2.*
- 32: gb_in3.*
- 33: gb_pl3.*
- 34: gb_pr4.*
- 35: em_ba1.*
- 36: em_ba2.*
- 37: em_btgl.*
- 38: em_btgt.*
- 39: em_btgt3.*
- 40: em_btgt4.*
- 41: em_btgt5.*
- 42: em_btgt6.*
- 43: em_btgt7.*

- 44: em_btgt8.*
- 45: em_btgt9.*
- 46: em_btgt10.*
- 47: em_hum3.*
- 48: em_hum4.*
- 49: em_hum5.*
- 50: em_hum6.*
- 51: gb_pr5.*
- 52: gb_pr6.*
- 53: gb_pr7.*
- 54: gb_btgl.*
- 55: gb_btgt2.*
- 56: gb_btgt3.*
- 57: gb_btgt4.*
- 58: gb_btgt5.*
- 59: gb_btgt6.*
- 60: gb_btgt7.*
- 61: gb_btgt8.*
- 62: gb_btgt9.*
- 63: gb_btgt10.*
- 64: gb_btgt11.*
- 65: gb_btgt12.*
- 66: gb_btgt13.*
- 67: gb_btgt14.*
- 68: gb_btgt15.*
- 69: gb_btgt16.*
- 70: gb_btgt17.*
- 71: gb_btgt18.*
- 72: gb_btgt19.*
- 73: gb_btgt20.*
- 74: gb_btgt21.*
- 75: gb_btgt22.*
- 76: gb_btgt23.*
- 77: gb_sts1.*
- 78: gb_sts2.*
- 79: gb_vil.*
- 80: gb_vil2.*
- 81: gb_pat1.*
- 82: gb_pat2.*
- 83: em_btgt0.*
- 84: gb_btgt24.*
- 85: gb_pr8.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1431.4	96.9	1481	2	LFU60594 U60594 Leptosira
2	1388.6	94.0	1450	2	LFA19243 Y19243 Leptosira
3	1295.6	87.7	1508	2	LIRRN16S X17547 Leptosira
4	1270.0	86.0	1374	2	L116SRDNY Z21634 L.inadai ge
5	1263.8	85.6	1483	29	WA5216S U12676 Leptosira
6	1263.8	85.6	1489	29	WA5216S U12677 Leptosira
7	1262.2	85.5	1513	2	LBUI12670 U12670 Leptosira
8	1261.0	85.4	1489	2	LBUI12669 U12669 Leptosira
9	1260.8	85.4	1516	2	LBUI12671 U12671 Leptosira
10	1254.6	84.9	1425	2	LB16SRDNP Z21630 L.borgpeter
11	1251.8	84.8	1486	29	WB4616S U12673 Leptosira
12	1246.6	84.4	1419	2	LK16SRDNP U12672 Leptosira
13	1239.8	83.9	1494	2	LSU12672 U12672 Leptosira
14	1238.2	83.8	1418	2	LM16SRDNY Z21648 L.meyeri ge
15	1232.8	83.5	1396	2	L116SRRN Z21649 L.santarosa
16	1230.0	83.3	1415	2	LS16SRDNY Z21637 L.weilli ge
17	1225.4	83.0	1410	2	LN16SRDNY Z21635 L.noquchii
18	1216.0	82.3	1424	2	LN16SRDNY M88721 Leptosira
19	1063.4	72.0	1523	2	LEPRR16SB Z98590 Leptosira
20	1031.8	69.9	1433	2	LBTM61993 Z21638 L.wolbachii
21	1027.6	69.6	1417	2	LB16SRDNY

22 1022.2 69.2 1439 2 LBTM31994 Z98588 Leptospira
23 1003.6 67.9 1451 1 AAL6SRDNP Z21629 A. ancona ge
24 1003.4 67.9 1321 2 LEPR16SA M71241 Leptospira
25 1003.2 67.9 1366 1 AF167353 AF167353 Leptospir
26 999.8 67.7 1442 2 LBTM31993 Z98587 Leptospira
27 995.8 67.4 1410 2 CC16SRDNP Z21631 C. canela ge
28 993 67.2 1398 2 LB16SRDNP Z21821 L. biflexa 1
29 991.8 67.1 1394 2 LB16SRDNP Z26969 L. biflex g
30 991.8 67.1 1394 2 LBTM21993 Z98592 Leptospira
31 991.6 67.1 1379 2 JJ16SRDNP M88719 Leptospira
32 985.8 66.7 1427 2 LBTM21994 Z21633 J. Jequitai
33 977.8 66.2 1466 2 LBTM11993 Z98589 Leptospira
34 972 65.8 1445 2 LB16SRDNP Z21632 L. illini ge
35 972 65.3 1413 2 LBTM61994 Z98591 Leptospira
36 964.4 62.0 1543 29 UEJ9479 AU009479 unculture
37 915.8 62.0 1385 2 LB16SRDNP Z26970 L. biflex g
38 884 59.9 1475 29 UEJ81652 U81652 Unidentifie
39 883.2 59.8 1475 29 UEJ81652 X77216 P. acidigall
40 874.4 59.2 1494 2 PA16SRDNP X75272 Clostridium
41 866.6 58.7 1494 2 CG16SRDNP Y17712 Malonomonas
42 866.6 58.7 1525 2 MRY17712 X72869 C. thermopal
43 864.2 58.5 1503 2 CT16SRDNP AB039334 Brevibaci
44 857.8 58.1 1532 1 AB039334 Z21636 L. parva gen
45 856.6 58.0 1379 2 LP16SRDNP

ALIGNMENTS

RESULT 1
LFU60594 1481 bp DNA BCT 10-SEP-1998
Leptospira fainei 16S ribosomal RNA gene, partial sequence.

U60594
U60594.1 GI:1408219
Leptospira fainei.
Leptospira fainei.
Bacteria; Spirochaetales; Leptospiraceae; Leptospira.
1 (bases 1 to 1481)
Perolat, P., Chappel, R.J., Adler, B., Baranton, G., Bulach, D.M.,
Billinghurst, M.L., Letocart, M., Merien, F. and Serrano, M.S.
Leptospira fainei sp. nov., isolated from pigs in Australia
Int. J. Syst. Bacteriol. 48 Pt 3, 851-858 (1998)

2 (bases 1 to 1481)
Adler, B., Chappel, R.J., Baranton, G., Bulach, D.M.,
Billinghurst, M.L., Letocart, M., Merien, F., Serrano, M.S. and
Perolat, P.
Direct Submission
Submitted (12-JUN-1996) Microbiology, Monash University, Wellington
Rd., Clayton, VIC 3168, Australia
Location/Qualifiers

source
1. .1481
/organism="Leptospira fainei"
/strain="Hurstbridge"
/db_xref="taxon:48782"
<1. .1481
/product="16S ribosomal RNA"

rRNA
BASE COUNT 391 a 335 c 439 g 314 t 2 others
ORIGIN

Query Match 96.9% Score 1431.4; DB 2; Length 1481;
Best Local Similarity 99.7%; Pred. No. 4; 1e-16;
Matches 1476; Conservative 0; Mismatches 1; Indels 4; Gaps 4;

Qy 1 gatcatgctcagaaactaacgctgcccgcgtctttaaacatgcgaatcgaagcgggtag 60
|||||

Db 1 GATCATGCTCAGAACTAACCGCTGGCGCGCTTAAACATGCAAGTCGAGCGGGTAG 60
|||||

Qy 61 caatcacctagcggcgaacgggtgagtaaacagt-ggtaalcttctccgagctgggata 119
|||||

Db 61 CAATACCTAGCGGCGAAGCGGTGAGTAACAGCTGGGTAATCTTCTCCGAGTCTGGGATA 120
Qy 120 acttccgaaaggaaagcctaataccggatagctctgttgatcacaaagattatgatagta 179
|||||
Db 121 ACTTCCGAAAGGAAAGCTAATACCGGATAGTCTGTGGATCAACAAGATTTCATAGGTA 180
|||||
Qy 180 aagattattcttggaatgagccgcgagcgtattagctagttggtgagtaagctc 239
|||||
Db 181 AAGATTATTCGTTGGAGATAGCCCGCGCGGATAGCTAGTTGGTGAGGTAATGGCTC 240
|||||
Qy 240 accaaggcagatcgttagcgcgctgaggggtgtccggcccaatgggaactgagaca 299
|||||
Db 241 ACCAAGGCGACATCGGTAGCGGCTGAGAGGGTGTCCGCCCACTAAGTAAGTGAACA 300
|||||
Qy 300 cgttcatactactcctcggagggcagcagtttaagaatcttctcaatgggggaaacccctga 359
|||||
Db 301 CGGTCCATACTCTCTACGGGAGGACAGTAAAGATCTTGTCAATGGGGGAAACCCCTGA 360
|||||
Qy 360 agcagcagcgcgctggaacaaagagctcttcgagttgtaaaagttcattaggcag-gaa 418
|||||
Db 361 AGCAGCGACCGCGGTGAACGAAGAGGCTTTCGGATTGTAAGTTTCATAGGCAGCAA 420
|||||
Qy 419 aaataagcagcaatgtgatgtgtacctgcctaaagcagccggtactacgtgcagca 478
|||||
Db 421 AATAAGCAGCAATGTGATGTGTACCTGCCTAAAGCACCAGCTAAGTACGTGCCAGCA 480
|||||
Qy 479 gccgcgtaatacgtatggtgcaagcgtttgttcgaatcattggcgtaaaaggtgcgta 538
|||||
Db 481 GCCCGGTATACGTATGGTGAAGCGTTTTCGGGAATCATTTGGCGGTAAAGGGTCCGTA 540
|||||
Qy 539 ggcggatttgtaagtcaggtgtgaaactgcgggtcacaacccgtggtgcactgaaac 598
|||||
Db 541 GCGCGATTGTAAGTCAGGTGTGAAACCTCGCGGCTCAACCCGTGCGCTTGAAC 600
|||||
Qy 599 tacaagctctggatttgggagagcgaagtggaaattccaggtgtagcgtgaaatgcgtag 658
|||||
Db 601 TACAAGTCTGGAGTTTGGGAGAGGCAAGTGAATCCAGGTGTAGCGGTGAATGCGTAG 660
|||||
Qy 659 atatctggaggaacaccagtgccgaagggcacttgcgtcctcaaaactgacgtgagga 718
|||||
Db 661 ATATCTGGAGGACACCAGTGGCGAAGGCGACTTGTGTGCTCAAAACTGACGCTGAGGA 720
|||||
Qy 719 cgaagcgtggtagtaaaacggattagataccccgggtaataccacccctaaacgtgttc 778
|||||
Db 721 CGAAAGCGTGGTAGTAACGGGATTAAGTACCCCGGTAAATCCACGCCCTTAAACGTTGTC 780
|||||
Qy 779 taccagttgtggggttttaacctcagtaacgaacctaaccgattaaagtagacgcct 838
|||||
Db 781 TACCAGTTGTTGGGGTTTAAACCTCAGTAACGAACCTAACGGATTAACTAGACCGCCT 840
|||||
Qy 839 ggggactatgctcgaagagtgaaactcaaggaattgacgggggttcgcacaaagcgtg 898
|||||
Db 841 GGGGACTATGCTCGCAAGAGTGAACCTCAAGAGGAATTGACGGGGGTCCGCACAAAGCGGTG 900
|||||
Qy 899 gacgatgtggttatttcgatacccccaccccaaacctcacctggcgttgacatggatctg 958
|||||
Db 901 GAGCATGTGCTTTAATTCGATGATACCCCAAAACCTCACCTGGGCTTGACATGGATCTG 960
|||||
Qy 959 aatcatgtagatataatgacgttcgggcagatttcacaggtgctgcatggttcttca 1018
|||||
Db 961 AATCATGTAGATATATGAGCTTCGGGCGAGATTCAGGTGTGATGTTGTGCTGCA 1020
|||||
Qy 1019 gctcgtgctggaatgttggttaagtcccgcaacgagcgaacccctatcgtatgttg 1078
|||||
Db 1021 GCTCGTGTGCTGAGATGTTGGTTAAGTCCCGCAACGAGCGCAACCCCTATCGTATGTTG 1080
|||||
Qy 1079 ctacc-ttaagtgggactggtacgaaactgcccgggtgacaaacccggaggaagcgggga 1137
|||||
Db 1081 CTACCATTAAGTTGGCACTCGTACGAAACTGCCGCTGACAAACCGGAGGAGCGGGGA 1140
|||||
Qy 1138 tgaagtcaaatcctcatggcctttatgtccagggccacacgctgctacagtggccgata 1197
|||||
Db 1141 TGACGTCAAATCCTCATGGCCCTTTATGTCCAGGGGCCACACAGCTGCTACATATGGCCGATA 1200
|||||


```
QY 1198 cagaaggtcgccaactcgcaagaggagctaatctctataaaagtcggtcccaagtcggtatt 1257
|||||
Db 1201 CAGAGGGTCGCCAACTCGCAAGAGGAGCTAATCTCTAAAGTCGGTCCCATCTCGGATT 1260

QY 1258 ggggtctgaactcgaccccaagaagtcggaatcgtagtaatcgcggaatcgatgcg 1317
|||||
Db 1261 GGGGTCGCAACTCGACCCCAATGAAGTCGGAATCGCTAGTAATCGCGGATCAGCATGCGG 1320

QY 1318 cggtaatactgtccggaccctgtacacacccgcgcgtcacaccacctgagtgggagca 1377
|||||
Db 1321 CGGTGAATACGTTCCGGACCTTGTACACACCGCCGCTCACACCACCTGAGTGGGAGCA 1380

QY 1378 cccgaagtggtcttctgtaaccgttaagagagacagactactaagtggaactcgtaaaagg 1437
|||||
Db 1381 CCCGAAGTGGTCTTCTGTTAAACCGTAAGGAGACAGACTACTAAGGTCAAACTCCTAAAGG 1440

QY 1438 ggtgaagtcgtatacaaggtta-ccgtaaatcgattcctgcag 1477
|||||
Db 1441 GGTGAAGTCGTAAACAGGTARCCGTAATCGATTCTCTGCAG 1481

RESULT 2
LOCUS LFA19243 1450 bp DNA BCT 08-NOV-1999
DEFINITION Leptospira fainei partial 16S rRNA gene, SSI 5402-98.
ACCESSION Y19243
VERSION Y19243.1 GI:6318184
KEYWORDS 16S ribosomal RNA; 16S rRNA gene.
SOURCE Leptospira fainei.
ORGANISM Leptospira fainei.
Bacteria; Spirochaetales; Leptospiraceae; Leptospira.
REFERENCE 1 (bases 1 to 1450)
AUTHORS Petersen,A.M., Krogfelt,K.A., Perolat,P., Boye,K. and
Schlichting,P.
TITLE Leptospira fainei serovar Hurstbridge isolated from two Danish
patients with Weil's syndrome
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1450)
AUTHORS Krogfelt,K.A.
TITLE Direct Submission
JOURNAL Submitted (05-NOV-1999) K.A. Krogfelt, Statens Serum Institut,
Dept. of Gastrointestinal Infections, Artillerivej 5, 2300
Copenhagen, DENMARK
FEATURES
Source
1..1450
Location/Qualifiers
/organism="Leptospira fainei"
/strain="SSI 5402-98"
/db_xref="taxon:48782"
<1..>1450
rRNA /gene="16S rRNA"
gene /product="16S ribosomal RNA"
1..1450
BASE COUNT 383 a 326 c 437 g 304 t
ORIGIN
Query Match 94.0%; Score 1388.6; DB 2; Length 1450;
Best Local Similarity 99.2%; Pred. No. 1.9e-15;
Matches 1427; Conservative 0; Mismatches 9; Indels 3; Gaps 3;

QY 24 ggcggcgctcttaaacatcgagtcgagcgggtagcaataccctagcgcgaacgggtg 83
|||||
Db 1 GCGCGCGCGCTCTTAAACATGCAAGTCGACGCGGGTAGCAATACCTAGCGCGCAACGGGTG 60

QY 84 agtaacacgt-ggtaatctctcgcagtcctgggataaacttccgaagaagaagtaata 142
|||||
Db 61 AGTAACACGTGGGTAACTCTCCGAGTCTGGGATAACTTTCCGAAGAAGAAAGCTAATA 120

QY 143 ccggatagtcctgttgatcaccaagatttagtgtaagatttatcttgagatgag 202
|||||
Db 121 CCGGATAGTCTCTACTGGATCACAGGATCTGTAGGTAAAGATTATATGCTTGGAGATGAG 180
```

```
QY 203 cccgcggccgatttagttagtgaggtaatcgctcaccaaggcgacgatcggtagccg 262
|||||
Db 181 CCCGCGCGCGATTAGTACTTGGTGAGGTAATGGCTCACCAAGGCGACCATCGTAGCCG 240

QY 263 gactgagagggtgctccggccacaatggaactgagacacggtccatactcctcaggagc 322
|||||
Db 241 GCCTGAGAGGGTGTCCGGCCACAATGGAACAGGTGAGACACGTCTACTACTCTACGGGAGC 300

QY 323 acgagttaagaatctctgtcaatgggggaaacccctgaagcagcgcgcgctgaacgaa 382
|||||
Db 301 AGCAGTTAAGATCTTGTCTCAATGGGGAAACCCCTGAAGCAGCAGCCGCGGTGAACGAA 360

QY 383 gaagttcttcggattgtataagttcattaggcagg-aaaaataagcagcaatgtgatg 441
|||||
Db 361 GAAGGTCTTCCGATTCTTAAAGTTCAATTAAGCAGGGAATAAAGCAGCAATGTGATGATG 420

QY 442 gtacctgcctaaagcacccgcgttaactacgtgccagcagccgcgggtaatacgtatg 501
|||||
Db 421 GTACCTGCCTAAAGCACCCGGCTTAACGTACGTGCACGAGCCGCGGTAAATACGTATG 480

QY 502 agcgtgttcggaatcattggcgtaaaagggtcgtagcgagattgttaagtcaggtg 561
|||||
Db 481 AGCGTTGTTCCGGAATCATTTGGCGGTAAAGGGTCCGTAGCGCGGATTGTAAAGTCAGG 540

QY 562 aaaaactgcgggctcaaccccgctggtgcacttgaaaactacaagctctggagtttgggag 621
|||||
Db 541 AAAAATCGCGGCTCAACCCGTGCGCTGCACCTTGAACACTACAAGTCTGGAGTTTGGGAG 600

QY 622 gcaagtgaattccagggtgtagcgtgaaatcgtagatatcttgaggaggaacaccagtgc 681
|||||
Db 601 GCAAGTGGAAATTCAGGTTAGCGGTGAAATCGTAGATATCTGGAGGAACACCAGTGGC 660

QY 682 gaagcgactctggtctcaaaactgacgtcagcgacaaagcgtggtgtagtaaacagg 741
|||||
Db 661 GAAGCGGACTTGTGCTCAAACTGACCTGAGGACCCGACCAAGCGGTGGTGTAAACGGG 720

QY 742 attagataccccgggttaatccacgcccctaaacggtgtgtctaccagttgttgggggtt 801
|||||
Db 721 ATTAGATACCCCGGTAATCCACGCCCTAAACGTTGTCTACCAGTTTGTGGGGTTTTAAC 780

QY 802 cctcagtaagaacacctaaggattaaagtagacgcgcctggggactatgctcgcaagatga 861
|||||
Db 781 CTTAGTAGAACAACTTACGGATTAAAGTAGACGCCCTGGGGACTATGCTCGCAAGAGTGA 840

QY 862 aactcaaaaggaattgacgggggtccgcacaaagcgggtggagcatgtgtttaaattcgat 921
|||||
Db 841 AACTCAAAAGGAATTGACGGGGTCCGCACAGGGGTGGAGCATGTGTTTAAATTCGATGA 900

QY 922 taccocaaaaacctcacctcgtggtgacatgacatgacatgacatgacatgacatgaccc 981
|||||
Db 901 TACGCGAAAAAACCCTCACCTGGGCTTGACATGGATCTGTAATCATGTAGAGATATATGAGCC 960

QY 982 ttccgggacagattcacaggtgctgcatggtgtcgtcagctcgtgctgtagatgttggt 1041
|||||
Db 961 TTCGGGCGATTTCACAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020

QY 1042 taagtcccgcaacgagcgcaacccctctatgattgtgtacc-ttaagttgggacactggt 1100
|||||
Db 1021 TAAGTCCCGCAACGAGCGCAACCCCTATCTGTATGTGCTTACCATTAAAGTTGGGCACTCGT 1080

QY 1101 acgaactgccggtgacaaaacccggaggaaggcggggtgacgtcacaatcctcatggccct 1160
|||||
Db 1081 ACGAAACTCCCGGTGACAAACCCGGAGAGGCGGGGATCACGTCAAAATCCTCATGSCCTT 1140

QY 1161 tatgtccaggggcacacacacgtctactacaatgcccatacagaggggtcgcaactcgcaaga 1220
|||||
Db 1141 TATGTCCAGGGGCACACACGTCTACAAATGGCCGATACAGAGGGTCCCAACTCGCAAGA 1200

QY 1221 gggagtaatctctataaagtcggtccagttcggatttgggttgcgactcgaactcgaacccatg 1280
|||||
Db 1201 GSGAGCTAATCTCTAAAGTCGGTCCAGTTCGGATTGGGGTCTGCACTGCACTGCACTG 1260
```

QY	1281	aagtcggaatcgtctagtaatcgcggatcagcatcccgcggtgaatcgttcccgacatt	1340
Db	1261	AAGTCGGAATCGCTAGTAATCGCGATCAGCATCCCGGCTGAATACGTTCCCGGACCTT	1320
QY	1341	gtacacacgcccgtcacaccactgagtggggagacccgaaagtgtctttgttaacgc	1400
Db	1321	GTACACACGCCCGCTCACACCCTGAGTGGGGAGCACCCGAAGTGTCTTTGTTAAACG	1380
QY	1401	taagagacagactactaagggtgaactcgtaaaggggtgaagtcgtacaaggatcc	1459
Db	1381	TAAAGAGACAGACTACTTAAGGTGAAACTCGTAAGGGGTGAAGTCGTACAAAGTAGC	1439
RESULT	3		
LIRRN16S		LIRRN16S 1508 bp DNA BCT 04-SEP-1991	
LOCUS		Leptospiira interrogans 16S ribosomal RNA gene.	
DEFINITION		X17547	
ACCESSION		X17547.1 GI:44008	
VERSION		16S ribosomal RNA; ribosomal RNA.	
KEYWORDS		Leptospiira interrogans.	
SOURCE		Leptospiira interrogans.	
ORGANISM		Bacteria; Spirochaetales; Leptospiraceae; Leptospiira.	
REFERENCE		1 (bases 1 to 1508)	
AUTHORS		Fukunaga, M.	
TITLE		Direct Submission	
JOURNAL		Submitted (22-NOV-1989) Fukunaga M., Faculty of Pharmacy and Pharmaceutical Science, Fukuyama University, 985 Higashimuracho, Fukuymashi, Hiroshimaken 729-0, Japan	
REFERENCE		2 (bases 1 to 1508)	
AUTHORS		Fukunaga, M., Horie, I., Okuzako, N. and Mifuchi, I.	
TITLE		Nucleotide sequence of a 16S rRNA gene for Leptospiira interrogans serovar canicola strain Moulton	
JOURNAL		Nucleic Acids Res. 18 (2), 366 (1990)	
MEDLINE		90221824	
COMMENT		Data kindly reviewed (23-FEB-1990) by Tukanaga M.	
FEATURES		Location/Qualifiers	
source		1..1508	
rRNA		/organism="Leptospiira interrogans"	
		/strain="serovar canicola strain Moulton"	
		/db_xref="taxon:173"	
		1..1508	
		/note="16S ribosomal RNA"	
BASE COUNT		393 a 342 c 456 g 317 t	
ORIGIN			
Query Match		87.78; Score 1295.6; DB 2; Length 1508;	
Best Local Similarity		94.78; Pred. No. 5.4e-14;	
Matches 1384; Conservative		0; Mismatches 74; Indels 4; Gaps	
QY	1	gatcattgctcagaactaacctgcgcgtcttaaacatcgaagtcagcgggtag	60
Db	15	GATCTCGGCTCAGACTAACGCTGCGCGCGCTTAAACATCAAGTCAAGCGGAGTAG	74
QY	61	caatacctagcggcggaacgggtgagtaaacgt-ggtaattcttcctccgagtcaggata	119
Db	75	CAATACTACGCGCGAACGGGTGAGTACACGTGGTAACTTCTCTGAGTCGGGATA	134
QY	120	acttccgaagaagactaaaccggatagtcctgttggatcacagaattgatagta	179
Db	135	ACTTTCGGAAGGAAGCTAATACTGGATGTCGCGAGAGATACAAAGATTTTCGCGTA	194
QY	180	aagattattgcttgagatgagcccgccgattagctagttggtgaggtaatggctc	239
Db	195	AAGATTTATTCTCGGAGATGAGCCGCGCTCCGATTACGTAGTGGTGAGGTAAGGCTC	254
QY	240	accaagcgacatcggtagccgctcgaggggtgtccggccaatggaaactgagaca	299
Db	255	ACCAAGGCGAGTCGCTAGCCGCGCTGAGAGGGGTGTCGCGCCAAATGGAACATGAGACA	314
QY	300	cggctccatactctacggggaggcaggttaagaattctgctcaatgggggaaacctga	359

Db	315	CGTCCATACTCTACGGGAGCGCAGCAGTTTAAAGAAATCTTGCTCAATGGGGGAA-CTTGA	373
Qy	360	agcagcgacgcgcgtgcaacgaagaaggtcttcgattgtaaagtctcattagca-ggaa	418
Db	374	AGCAGGAGCGCGGTGAACGATGAAGGCTCTTCGGATTGTAAGATTGTAAGACGAGGAA	433
Qy	419	aataaagcagaatgtgatgtgacctgctctaaagcaccgcgctaactacgtgccagca	478
Db	434	AAATACGACGAATGATGATGGTAGGTACCTTGGCTTAAGCACCAGGCTAACTACGTCCAGCA	493
Qy	479	gccgcggttaacgtatgtgcaagcgttgttcgggaatcattggcgctaaaggggtgcgta	538
Db	494	GCCCGGTAAATACGATATGTGTCAAGCGTTTTCGGAATCAITTTGGCGCTTAAGGCGTGA	553
Qy	539	ggcggaatttgtaagtcagggtgtgaaactgcgggctcaaccgcgtgacctgcacttgaac	598
Db	554	GGCGACATGTAACTGACAGTGTGAAACATGCGGGCTCAACTCGCAGCCTGCACTTGAAC	613
Qy	599	tacaagtcgagtttgagagagcgaagtggaaattccagtgtagcgtgtaaaatgcgtag	658
Db	614	TATGTGCTGGNGTTTGGAGAGGCAAGTGGAAATTCACAGGTACGGGTGAATATCGGTAG	673
Qy	659	atatctggaggaaacaccagtagtgcgaagcgacttgtcgtcctaaaaactgaagctgagga	718
Db	674	ATATCTGAGAGAACACCAGTGGCGAAGCGCACTTGTGGCTAAAACTGACGCTGAGGCA	733
Qy	719	cgaagcgtgggtagttaaacgggattagatacccggtaatccagccctaaccgtgtgc	778
Db	734	CGAAGCGTGGTGTGTAACGGATTAAGTACCCCGGTAATCCACGCCCTTAACGTTGCTC	793
Qy	779	taccagttgtgggggttttaacctcagtaacgaacctcaacggattaagttagaccgcct	838
Db	794	TACAGTTGTGGGGGTGTTTAACTCTAGTAACGAACCTAACGGATTAAGTAGACGCCCT	853
Qy	839	gggactatgctgcgaagagtgaactcaaggaattgacgggggtccgcacaagcgggtg	898
Db	854	GGGACTATGCTCGCAAGAGTCAAACTCAAAGGAATTACGGGGGTCCGCACAAGCGGTG	913
Qy	899	gacatggtttaatttcgatgatatacccaaaaacctcacctggcctgacatgagctg	958
Db	914	GAGCATGTGTTTAATTCGATGATACGCCAAAACCTCACCTAGGCTTGACATGGAGTGG	973
Qy	959	aatcatgtagagatatagacctcgccagattcacagggtgcgtgcagtggtctgcgca	1018
Db	974	AATCATGTAGAGATACATGAGCCTTGGGCCGCTTCACAGGTGCTGATGTTGCTGCTCA	1033
Qy	1019	gctggtgctgtagatgttggttgaagtccgcgaagcgaacccctatctgtagttg	1078
Db	1034	GCTCGTGTGCTGAGATGTGGGTTTAACTCCGCAACGAGCGCAACCTCACCTTATGTTG	1093
Qy	1079	cta-ccctaagttagcactggtacgaaactgcggtgcaaacccgaggaagcgcgga	1137
Db	1094	CCATCATTCAGTTGGGCACTCGTAAGGAACCTGCGGTGACAAACCGAGGAAGCGGGGA	1153
Qy	1138	tgagtcgaactcctcatggcctttatgtccaggccacacacgtgtctaaatggcggata	1197
Db	1154	TGACGTCAAAATCCTCATGGCCTTTATGTCTTAGGCGACACACGCTGTACAAATGGCGGTA	1213
Qy	1198	cagaggtgcgaactcgcaagaggagactaatctctaaagtcggtcccgatcgagatt	1257
Db	1214	CAAAGGTAGCCAACTCGCGAGGGGAGCTAATCTCAAAAATCCGGTCCCGATTCGGATT	1273
Qy	1258	ggggctctgaactcgaccccatgaagtcggaatcgtagtaatcgcgagatcagcatgcgcg	1317
Db	1274	GGAGTCTGCAACTCGACTCCATGAAGTCGGATCCGTAGTAAATCGCGGATCAGCATGCCG	1333
Qy	1318	cgttgaaatagttcccggaacctgttacacacgcgcgcgtcacacacactgagtgaggagca	1377
Db	1334	CGGTGATATGTTCCCGGACCTTGTACACACCGCGGTACACCCACTGAGTGGGAGCA	1393
Qy	1378	cccgaagtggctcttgtaaccgttaaggagacagactactaaggtaaacctcgtaaaagg	1437
Db	1394	CCCGAAGTGGTCTTTTGCCACCCGCAAGGAGCAGACTACTAAGTGAACCTCGTGAAGGG	1453

```
Qy 1438 ggtgaagtcgttaacaaagtacc 1459
|||||
Db 1454 ggtgaagtcgttaacaaagtacc 1475

RESULT 4
L16SRDNY L16SRDNY 1374 bp DNA BCT 02-DEC-1993
LOCUS L.inadai gene for 16S ribosomal RNA (partial).
ACCESSION Z21634
VERSION Z21634.1 GI:433581
KEYWORDS 16S ribosomal RNA.
SOURCE Leptospira inadai.
ORGANISM Bacteria; Spirochaetales; Leptospiaceae; Leptospira.
REFERENCE 1 (bases 1 to 1374)
AUTHORS Hooke, J.V.
TITLE Phylogeny of Leptospiaceae and related Spirochaetes
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1374)
AUTHORS Hooke, J.V.
TITLE Direct Submission
JOURNAL Submitted (09-FEB-1993) Hooke J.V., Public Health Laboratory
Service, Leptospira Reference Laboratory, Stonebow Road, Hereford,
Herefordshire, United Kingdom, HK1 2ER
LOCATION/Qualifiers
FEATURES
source
1..1374
/organism="Leptospira inadai"
/strain="LYME"
/db_xref="taxon:29506"
<1..>1374

rrna
/product="16S ribosomal RNA"
358 a 311 c 413 g 289 t 3 others

BASE COUNT
ORIGIN

Query Match 86.08; Score 1270; DB 2; Length 1374;
Best Local Similarity 96.58; Pred. No. 1.5e-13;
Matches 1327; Conservative 2; Mismatches 43; Indels 3; Gaps 3;

Qy 16 ctaacgctggcgccgtcttaacatcgaagtcgagcggtgagcactacacgagcg 75
|||||
Db 1 STARCGTTGGCGGCGCTCTTAACATCATCAAGTCGAGCGGGGTAGCAATACCTAGCGCG 60

Qy 76 aacgggtgaagtaacacgt-ggtaattcttcctccgagtcggtggataaactttccgaagaa 134
|||||
Db 61 AACGGTCAGTAACACGTTGGTAACTTCTCCGAGTCTGGGATAACTTCCGAAGGAA 120

Qy 135 agctaataccggatagtcctgttgatcacagaatttatgaggttaagattattgcttg 194
|||||
Db 121 AGCTAATAACCGGATAGTCTACTGATCAGAGGATCTGATAGGTAAAGATTATTGCTTG 180

Qy 195 gagatgagccgcccagcttagctagttgtgaggttaagttgctcaccagcagcagtc 254
|||||
Db 181 GAGATGACCCCGCCGCCATAGTAGTTGGTGGTAAAGCTCACCAGCGACGATC 240

Qy 255 ggtagccgctgagaggggttcggccacaaatggaactgagacacggtccatactccta 314
|||||
Db 241 GGTAGCCGCTGAGAGGGTGTCCGGCCACAATGGAACGTGAGACACGGTCCATCTCCTA 300

Qy 315 cgggagcagcaggttaagaattctgtcctaattgggggaacacctgaagcagcagccgcg 374
|||||
Db 301 CGGGAGCAGCAGGTATGAATCTTGTCTCAATGGGGGAACCTGAAAGCAGCGACGCCCG 360

Qy 375 tgaacgaagaaggtcttcgaggttgaagttcattagga-ggaataaagcagcaatg 433
|||||
Db 361 TCAACCAAGAGAGGTCTTCGGATTGTAAGCTTCAATTAACGACGGGAAATTAAGCAATG 420

Qy 434 tgatgatggtacgttcctaaagcagcggcctaactacgtgacgagcgcggtatacgt 493
|||||
Db 421 TCATCATGTACCTGCCTAAAGCACCAGCGCTAACACGTGACGTGCCAGCAGCGCGGTAATACGT 480
```

RESULT 5

```
WA4516S
LOCUS Leptospira weilii 1483 bp DNA BCT 18-FEB-1995
DEFINITION Leptospira weilii Cellidoni (ATCC 43285) 16S rRNA gene, partial
sequence.
ACCESSION U12676
VERSION U12676.1 GI:558937
KEYWORDS Leptospira weilii.
ORGANISM Leptospira weilii.
REFERENCE 1 (bases 1 to 1480) Bacteria; Spirochaetales; Leptospiaceae; Leptospira.
```

Ralph, D. and McClelland, M.
Phylogenetic evidence for horizontal transfer of an intervening
sequence between species in a spirochete genus
J. Bacteriol. 176 (19), 5982-5987 (1994)
95014031
2 (bases 1 to 1483)
McClelland, M.
Direct Submission
Submitted (25-JUL-1994) Michael McClelland, California Institute of
Biological Research, 11099 North Torrey Pines Road, La Jolla, CA
92037, USA
Location/Qualifiers
1. 1483
/organism="Leptospira wellii"
/strain="serogroup Celledoni, serovar celledoni, strain
Celledoni (ATCC 43285)"
/isolate="WA45"
/db_xref="taxon:28184"
<1. 1483
/product="16S rRNA"
BASE COUNT 382 a 331 c 452 g 315 t 3 others
ORIGIN

Query Match 85.68; Score 1263.8; DB 29; Length 1483;
Best Local Similarity 94.18; Pred. No. 1.7e-13;
Matches 1376; Conservative 0; Mismatches 80; Indels 6; Gaps 6;

QY 1 gacatggctgaactaacgctggcgccgctcttaaacatgcaagtcagcgggtag 60
DB 10 GATCTGGCTCAGAACTAACCGTGGCGCGGCTCTTAACATGCAAGTCAAGCGCAGTAG 69
QY 61 caatactagcgccgaacggtgagtaacacgt-ggtaattctctcgaactcggata 119
DB 70 CAATACTAGCGCGCAACGGGTGAGTAACACGTGGGTAACTCTCTCCGAGCTCGGGATA 129
QY 120 actttccaaagaaagctaataccgagatagctctgttggatcacaaagattgtagta 179
DB 130 ACTTTCGGAAGAGGGAGCTAATCTGATGGTGGTCCGAGAGGTCTATGATGTTTCGGGTA 189
QY 180 aagattattgttgagatgagcccgccgagcagtagctagctagtgtaggtaattgctc 239
DB 190 AAGATTATTTGCTCGGAGCTGAGCCCGCCCGCATAGCTAGTTGGTAGGTAATGGCTC 249
QY 240 accaagcgacatcgttagcgccctgaggggtgctccgcccacaaatggaactgagaca 299
DB 250 ACCAAGCGCAGATCGGTAGCCCGCTGAGAGGGTGTTCGGCCACAATGGAACCTGAGACA 309
QY 300 cggctcatactctacgaggagcagcagtttaagaatcttctcaatgggggaaacccctga 359
DB 310 CGGTCCATACT-CTACGGGAGGAGCAGTAAAGAACTTTGCTCAATGGGGGAAACCCCTGA 368
QY 360 agcagcaccgctggaacgaagaaggctctcgagattgtaagttcattaggca-ggaa 418
DB 369 AGCAGCACGCCCGGTGAACGATGAAGCTTCGGATTTCGGATTGAAGTTCAATAAGCAGGAA 428
QY 419 aataaagcagaatgtgatgtacctgcctaaagcaccggctaacacagtcagtcagca 478
DB 429 AATAAGCAGGATGTGATGATGCTACCTGCCTAAAGCAGCGGTAACTACGTGCCAGCA 488
QY 479 gcccggttaatacgtatggtgcaagcgtttctcggaatcattggcgtaaaaggtgcgta 538
DB 489 GCCCGGTGAATACGTATGTTGCAAGCGCTGTTCCGAATCAATGGCGTAAAGGGTCCGTA 548
QY 539 gccgatttgaactcaggtgtgaaactcgggctcaacccctggcctgacctgaaac 598
DB 549 GCCGGCATGTAACTCAGGTGTGAACACTCGGGCTCAACTCGCAGCGCTGCACCTTGAAC 608
QY 599 tacaagctgaggtttgggagaggaagtggaattccaggttagcggtgaaatcgtag 658
DB 609 TATGTCTGGAGTTGGGAGAGGCAAGTGAATTCAGGTAGCGGTGAAATGGGTAG 668
QY 659 atatctgaggaacaccagctggcgaaggcacttgcgtgctcaaacactgacgctgaggca 718

DB 669 ATATCTGGAGAACACACAGTGGCGAAGCGACTTGTGTGGCTAAAACCTGACGCTGAGGCA 728
QY 719 cgaagcgtgggtagtagtaaaacggttagatatacccccgtataatccacgcctaaacgttgct 778
DB 729 CCAAAGCGTGGGTAGTAGAACGGATTAGATACCCCGGTAAATCCACGCCCTTAACAGCTGTC 788
QY 779 taccagttgtgggggttttaacccctcagtaacgaacctaacggattaaatagacgcct 838
DB 789 TACCAGTTGTTGGGGTTTAAACCTCAGTAACGAA-CTAACGGATTAAAGTAGACCGCT 847
QY 839 ggggactatgctcgaagagtgaactcaaaaggatgaacgggggtccgcacaacggtg 898
DB 848 GGGGACTATGCTCGCAGAGTGAACACTCAAAGGAATTGACGGGGTNNNCAACAGCGGTG 907
QY 899 gacatggttttaattcgtatgatacccccacaaacctcaacctgggcttgacatggatctg 958
DB 908 GAGCATGTGGTTTAAATTCGATGATACGCGAAGAAACCTCACCTAGCTTGACATGGAGTG 967
QY 959 aatcatgtagagatatagccttcggcgagattcacaggtgctgcatggttctcgtca 1018
DB 968 AATTATGTAGAGATACATAGAG-CTTCGGGCGCTTCACAGGTGCTGATGTTGTGTCGA 1026
QY 1019 gctcgtgctgtagatgttggttaagtcccgcaacgagcgcaacccctatcgtatgtt 1078
DB 1027 GCTCGTGTGATGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCCTCACCTTATGTTG 1086
QY 1079 cta-cccttaagtgtggcactggtacgaacctgcccgtgacaaacggaggagcgggga 1137
DB 1087 CCATCATTTTGTGGGCATCTCGTAGGAAGAACTCCCGTGCACAAACCGAGAGCGGGGA 1146
QY 1138 tgacgtcaaatctcctatggcctttatgtccaggggcacacagctgctacaatggcgata 1197
DB 1147 TGACGTCAAAATCTCATATGCTTTATGTCTAGGGCAACACAGTGTCTACATATGGCCGTA 1206
QY 1198 caagagtcgcaactcgcgaagggagcgtatctctaaagtcggtcccgatctcgatt 1257
DB 1207 CAAAGGATGCCAACTCGCGAGGGGAGCTAATCTCAAAAACCGGCTCCAGTTCGGATT 1266
QY 1258 ggggtctgcaactcgaccctcgaagtcggaatcgtagtaatcggtatcggtacgacatccg 1317
DB 1267 GGAGTCTGCAACTCGACTCCATGAAGTCGGAATCGTAGTAATCGCGGATCAGATGCCG 1326
QY 1318 cgtggaatacgttcccgaccctgtacacaccccgctcacaccctcgtagtggtggagca 1377
DB 1327 CGGTGAATAGTTCCTCCGAGCTTGTACACACCGCGTCACACCACTGAGTGGGAGCA 1386
QY 1378 ccggaagtgtctttgttaaccgttaaggagacagactactaaagtgaaactcgtaaagg 1437
DB 1387 CCCGAAGTGGTCTTTGCAACCGTAAAGAAAGCAGACTTACTAAGGTGAAACTCGTGAAGG 1446
QY 1438 ggtgaagctgtaacaagggtacc 1459
DB 1447 GGTGAAGTCGTAAACAGGTACC 1468

RESULT 6

WA5216S 1489 bp DNA BCT 18-FEB-1995
LOCUS Leptospira wellii Worsfold 16S rRNA gene, partial sequence.
DEFINITION U12677
ACCESSION U12677.1 GI:5589938
VERSION U12677.1
KEYWORDS
SOURCE Leptospira wellii.
ORGANISM Leptospira wellii
Bacteria; Spirochaetales; Leptospiraceae; Leptospira.
REFERENCE 1 (bases 1 to 1480)
AUTHORS Ralph, D. and McClelland, M.
TITLE Phylogenetic evidence for horizontal transfer of an intervening
sequence between species in a spirochete genus
J. Bacteriol. (1994) In press
REFERENCE 2 (bases 1 to 1489)
AUTHORS McClelland, M.

/organism="Leptospira borgpetersenii"
/strain="serovar hardjo, strain Hardjo bovis/Sponselee"
/isolate="BD50"
/db_xref="taxon:174"
<1..1513
/product="16S rRNA"

BASE COUNT 385 a 346 c 464 g 315 t 3 others
ORIGIN

Query Match 85.5%; Score 1262.2; DB 2: Length 1513;
Best Local Similarity 94.0%; Pred. No. 1.8e-13;
Matches 1375; Conservative 0; Mismatches 81; Indels 6; Gaps 6;

```
Qy 1 gatcatggctcagaactaaagctggcgccgctcttaaacatcaagtcagcgggtag 60
Db 40 GATCCTGGCTCAGAACTAACCTGGCGCGCTCTTAACATGCAAGTCAAGCGGAGTAG 99
Qy 61 caatacctagcggcgaacgggtgagtaacagt-ggtaactcttctcccgagctcgggata 119
Db 100 CAATACTCAGCGCGGACAGGGGTGAGTACAGCTGGGTAACTTCTCCGAGCTGGGATA 159
Qy 120 actttccgaaggaaagctaataccgcatagtcctgttggatcacagaatttgatagta 179
Db 160 ACTTTCCAAAGGGGAGCTAATCTGATAGTCCCGAGAGCTCATAGGATTTTTCGGGTA 219
Qy 180 aagatttattcttgagatagacccgcgcgagattagctagttggtgagtaatgctc 239
Db 220 AAGATTTATGCTCGGCGCTGAGCCCGCGCCGATAGTACTGTTGGTGGAGTAAATGCTC 279
Qy 240 accaagggcagcatcggtagcggcctgagaggggtgtccggccacaatggaaactgagaca 299
Db 280 ACCAAGGCGACGATCGGTAGCGGCTGAGAGGGTGTTCGGCCACAATGGAACCTGAGACA 339
Qy 300 cggctcactctcagggagggcagcagtaagaatcttgctcaatgggggaaacctga 359
Db 340 CGGTCCATACT-CTACGGGAGCGCAGCAGTTAAGAACTTCGTAATGGGGGAAACCTTGA 398
Qy 360 agcagcgacgcgcgtgaacgaagagctctcgattgtaaaagttcattaggca-ggaa 418
Db 399 ACCAGCGACGCGCGCTGAACGATGAAGCTTCGCGATTGTAAGTTCAATAAGCAGGGAA 458
Qy 419 aaataagcagcaatgtgatgtgtacctgcttaagcaccggctaaactagtcgccaaga 478
Db 459 AAATAAGCAGCGATGTGATGTGATGTACCTGCTTAAGCACCAGCGCTAACCTAGTCCAGCA 518
Qy 479 gccgcgtaatacgtatgttgaacgctgttccggaatcattggcgtaaaaggggtgcga 538
Db 519 GCGCGGTAATACGTATGTGTCAACGGTTGTTCCGAAATCATTTGGCGGTAAAGGTTGCGTA 578
Qy 539 ggcgatttgaagtcaggtgaaactgcgggctcaaccgctgagcctgcactgaaac 598
Db 579 GCGGACATGTAACTCAGTGTGAAACTGCGGGCTCAACTCGCAGCTGCACCTTGAAC 638
Qy 599 tacaagctcggagtttggagaggaagtggaattccagggttagcgggtgaaatcgtag 658
Db 639 TATGTCTCTGGAGTTTGGGAGAGGAAGTGAATTCAGGTGTAGCGGTGAATCGGTAG 698
Qy 659 atatctggaggaacaccagtgccgaagcgacttgcgtcctcaaaactgaagcctgagga 718
Db 699 ATATCTGGAGGAACACCACTGCGGAGGCGCACTTCTGCGCTTAAACTGACGCTCAGGCA 758
Qy 719 cgaagcgtgggttagtaaacgggattagatacccccgggtaattccagccctaaccattgtc 778
Db 759 CGAAAGCGTGGGTAGTGAACCGGATAGATACCCCGGTAAATCCAGCGCTTAAACCTTCTC 818
Qy 779 taccagtttgggggttttaacctcagtaacgaacctcaagcagtaagtagcgcct 838
Db 819 TACCAGTTGTGGGGGTTTTTAACTCTCAGTACGAA-CTACGGATTAAGTAGACCGCT 877
Qy 839 ggggactatgctcgaagagtgaaactcaaaaggaattgacgggggtccgcacaaagcgtg 898
Db 878 GGGGACTATGCTCGCAAGAGTGAACCTCAAGGAATTTGACGGGGTNNNCACAAAGCGGTG 937
```

```
Qy 899 gagcatgtgttttaattcatgatacccccaaaacacctcagctggccttgacatggtctg 958
Db 938 GAGCATGTGTTTAAATTCATGATAGCGCAAAACCTCAGCTAGCTTGCATGGAGTGG 997
Qy 959 aatcatgtagagatatagagccttcggcgagattcaacagggtgctgcatggttgctgta 1018
Db 998 AATCATGTAGAGATACATAGAG-CTTCGGGCGCTTTCACAGGTCCTGCATGGTTGCTGCA 1056
Qy 1019 gctcgtgctgagatgttgggttaagtcctcccaagcagcgcaacccctatcgatgttg 1078
Db 1057 GCTCGTGTCTGAGATGTTGGTTAACTCCCGCAAGCAGCGCAACCTCACCTTATGTTG 1116
Qy 1079 cta-cttaagttgggcaactggtaacaaaactccggtgacaaacccggaggaagcgggga 1137
Db 1117 CCATCATTCAGTTGGCACTCGTAAGGAATCGCGGTGACAAACCGGAGGAGCGGGGA 1176
Qy 1138 tgcgtcaaatcctcagtcgctttatgtccaggggcacaacagctgctacaatggccgata 1197
Db 1177 TGAGCTCAATCTCATGGCTTTTATGCTAGGGCAACACGCTGTACAAATGGCCGGTA 1236
Qy 1198 cagaggtcgcgaactcgcgaagggaggtcctaactctctaaaagtcggtcccgatcgatt 1257
Db 1237 CAAAGAGTACCCAACTCGCGAGGGGAGGTAACTCTCAAAAGCGCGTCCAGTTCGGATT 1296
Qy 1258 ggggtctgcaactcgaccccatgaagtcggaaatcgctagtaatcgcggtatcagcatgcgg 1317
Db 1297 GGAGTCTGCAACTCGACTCCATGAAGTCGGAATCGCTAGTAATCGCGGATCAGCATGCCG 1356
Qy 1318 cgggtgaatcgttcccgacccctgtacacaccccgctcacaccactagtgaggagca 1377
Db 1357 CGGTGAATAGTTCCCGGACCTTTGTACACACCCCGCTCACACCTAGTGGGAGCA 1416
Qy 1378 ccgaaagtcgtcttggtaaaccgttaagagagacagactactaaagtgaaactcgtaaaagg 1437
Db 1417 CCCGAAGTGTCTTTGGCAACCCGCAAGGAGCAGACTACTAAGTGAAACTCGTGAAGG 1476
Qy 1438 ggtgaagtcgtaacaaggatcc 1459
Db 1477 GGTGAAGTCGTAACAAGGTAGC 1498
```

RESULT 8

LBUI2669 1489 bp DNA BCT 18-FEB-1995
LOCUS Leptospira borgpetersenii 1627 Burgas 16S rRNA gene, partial
SEQUENCE
U12669
U12669.1 GI:558928
SOURCE Leptospira borgpetersenii.
ORGANISM Leptospira borgpetersenii
Bacteria; Spirochaetales; Leptospiraceae; Leptospira.
REFERENCE 1 (bases 1 to 1489)
Ralph, D. and McClelland, M.
Phylogenetic evidence for horizontal transfer of an intervening
sequence between species in a spirochete genus
J. Bacteriol. 176 (19), 5982-5987 (1994)
MEDLINE 95014031
REFERENCE 2 (bases 1 to 1489)
McClelland, M.
Direct Submission
JOURNAL Submitted (25-JUL-1994) Michael McClelland, California Institute of
Biological Research, 11099 North Torrey Pines Road, La Jolla, CA
92037, USA
FEATURES
source
1. 1489
/organism="Leptospira borgpetersenii"
/strain="serovar balcanica, strain 1627 Burgas"
/isolate="BD30"
/db_xref="taxon:174"
<1..1489
/product="16S rRNA"
rRNA

BASE COUNT	385	a	330	c	454	g	311	t	9	others
ORIGIN										
Query Match			85.4%		Score 1261;	DB 2;	Length 1489;			
Best Local Similarity			93.8%		Pred. No. 1.9e-13;					
Matches 1372;			Conservative		Mismatches 0;		Indels 6;			Gaps
Qy	1	gatcatggctcagaactaacgctg	gcgcgcctcttaaacatgcaagtcgagcgggtag	60						
Db	16	GATCCTGGCTCAGACTAACGCT	GGCGGCGCTCTTAACATGCAAGTCACGCGAGTAG	75						
Qy	61	caatactagcggcgcaacgggtg	agtaacacgt-ggtaatttctctccgagctcgggata	119						
Db	76	CAATACTACGCGCAACGGTGAGT	AAACACGTGGGTAACTCTCTCCGAGCTCTGGGATA	135						
Qy	120	acittccgaaagaaagctaatacc	gcgagtagtccctgttgatcacagaattgataagta	179						
Db	136	ACITTCGAAAGGGAGCTAAATAC	TGGATAGTCCCGACAGGTCTATAGGATTTTTCGGGTA	199						
Qy	180	aagatttattgcttgagatgagcc	gcgcgcgattagctagttggtgaggtaaatggctc	239						
Db	196	AAGATTATTGCTCGGAGATGAGC	CGCGCCGATTAGTCTAGTTGGTGAGGTAAATGGCTC	255						
Qy	240	accaagcgacgacgtcggtgacg	cgccgtgagaggggtgccggccacaaatggaaactgagaca	299						
Db	256	ACCAAGCGACGATCGGTAGCCGG	CTGAGAGGGGTTCGGCCACAAATGGAACCTGAGACA	315						
Qy	300	cgtccatactcctacggagcagac	agattagaaattctcaatggggaaacccctga	359						
Db	316	CGGTCCCACT-CTACGGAGAGC	AGCAGTTAGAAATCTTGCTCAATGGGGGAAACCCCTGA	374						
Qy	360	agcagcgacgcgctgaaacgaag	gtcttcggattgtaaaagtctcataggca-ggaa	418						
Db	375	AGCAGCGACGCCGCTGAACGAT	GAAAGTCTTCGGATGTTAAAGTTCAATAGCAGGGAA	434						
Qy	419	aaataagcagcaatgtgatgtg	ttacgtgcctaaagcaccgggtacactacgtgccagca	478						
Db	435	AAATAAGCAGCATGTGATGATG	TACCTACCTTAAAGACACGGCTTAACCTAGCTGCCAGCA	494						
Qy	479	gcgcggtaatacgtatggtgca	gcgttggttcggaatcattgggcgtlaaagggtgcgta	538						
Db	495	CCCGCGGTAAATAGTATGTGTG	CAACGGTTTTCGGAAATCATTTGGGCGTTAAAGGTTGCGTA	554						
Qy	539	ggcggattgttaagtcaggtgtg	aaactgcgggctcaaccgctggcctgcacttgaaac	598						
Db	555	GGCGGACATGTAGTCAGGTGTG	AAACATGCGGGCTCAACTCGCAGCTGCCTTGAAC	614						
Qy	599	tacaagctggagtttgggagagc	aaatgggaaattccagggtgtagcggctgaaatgcgtag	658						
Db	615	TATATGCTCGAGTTTGGGAGAG	CGCAAGTGAATTCACAGTGTAGCGGTGAATAGCGGTAG	674						
Qy	659	atatctggaggaacacagtggc	gaaggcgactgctggctcaaaactgacgctgagga	718						
Db	675	ATATCTGGAGGAACACGATGG	CGGAAGGCGATTGCTGGCTTAAACCTGACGCTGAGGCA	734						
Qy	719	cgaagcgctgggtagtaaacg	ggattagatcccccggttaatccacgcccctaaacgcttgct	778						
Db	735	CGAAAGCGTGGTAGTCACG	GGAATAGATACCCCGGTAAATCCACGCCCTTAACAGTTGTC	794						
Qy	779	taccagttgttgggggtttta	acccttcagtaacgaacctaaaggataagtagaccgct	838						
Db	795	TACCAGTTGTTGGGGGTTTTA	ACCCCTCAGTAACGAA-CTAACGGATTAAGTAGACCGCCT	853						
Qy	839	ggggactatgctcgaaagagt	gaaactcaaaagaaattgacgggggtlccgcacaagcgggtg	898						
Db	854	GGGGACTATGCTCGCAAGAGT	GAAGCTCAAAAGGAATTCACGGGGGTNNNCACAGCGCGTG	913						
Qy	899	gagcatgtggtttaattcgat	gatacccccaaaaacctcacctgggcttgacatggatctg	958						
Db	914	GAGCATGTGGTTTAATTC	AGTATACCGCGAAACACCTCACCTAGGCTTTGACATGGAGTGG	973						
Qy	959	aatcatgtagagatatatga	gccttcgggcagatattcacaggtgctgtgcgtgtgtctgca	1018						

Db	974	AATCATGTACAGATACATGAG-C	TTTTGGGCCCGCTTTACAGAGTGC	TGTGATGCTTCGTCTCA	1032
Qy	1019	qclcggtcgtgagatggttgaagtc	ccgcacacagcgcacacccctatc	gtatattg	1078
Db	1033	CTCTGCTGCTGAGATGTTGGGT	TAAAGTCCCCCAAGGAGCAACCC	TACACCTTATGTTG	1092
Qy	1079	cta-ccctaaagttgggcactggt	tcagaaactgccggtgacaaaaccg	gaggaagcgcgga	1137
Db	1093	CCATCATTCAGTTGGGCACTCG	TAAAGAACTGCGGTGACAACTG	GAGGAAGCGGGGA	1152
Qy	1138	tgacgtcaaatccctatgcttgc	ttatgtccaggggccacacgtgct	tacaaatggccgata	1197
Db	1153	TGACGTCAAHNNNNNATGGCC	TTTATGTTAGGGCAACACACGCT	GCTACAAATGGCCGGTA	1212
Qy	1198	cagaggtgcaccaactcgacaag	agagctaattcctaaaagtcggt	ccgtcccgatccgatt	1257
Db	1213	CAAGGGTAGCCAACTCGGAGG	GGAGCTAATCTCAAAAGCCGGT	CCCCAGTTCGGATT	1272
Qy	1258	ggggtctgcaactcgaccccatg	aagtcggaatcgctagtaatcg	cgatcagcatgccg	1317
Db	1273	GGAGTCTGCAACTCGACTCC	ATGAAGTCGGAATGCTAGTAAT	CGCGGATCAGCATCCG	1332
Qy	1318	cgggtgaatacgttcccgacctt	gtacacacgcgcgcgtcaacac	ctgagtcggggagca	1377
Db	1333	CGGTGAATAGCTTTCCCGGAC	CTTGTACACACGCCCGCTCAC	ACCACCTTGAGTGGGGAGCA	1392
Qy	1378	ccggaagtggctcttcttaacg	taagcagagacagactactaag	tgaacctcgtaaagg	1437
Db	1393	CCCGAAGTGGTCTTTGCCAAC	CCCAAGGAGGAGCACTACTAAG	GGTGAACCTCGTGAAGGG	1452
Qy	1438	ggtagaagtcgtaacaaggtag	c 1459		
Db	1453	GGTGAAGTCGTAACAAGGT	AGC 1474		
RESULT 9					
LNUI2671					
LOCUS					
DEFINITION					
ACCESSION					
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
FEATURES					
source					
1. .1516					
/organism="Leptospira noguchii"					
/strain="serogroup Autumnalis, subserogroup Fort Bragg"					
/isolate="NE36"					
/db_xref="taxon:28182"					
<1. .1516					
/product="16S rRNA"					
BASE COUNT					
ORIGIN					
392 a 342 c 462 g 314 t 6 others					
Query Match 85.4%; Score 1260.8; DB 2; Length 1516;					
Best Local Similarity 93.9%; Pred. No. 1.9e-13;					

Matches 1373; Conservative 0; Mismatches 83; Indels 6; Gaps 6;

```
Qy 1 gatcatggtcagaactaacgctgctgctgtctttaaacaatcgaagtcgagcgggtag 60
Db 43 GATCTGGCTCAGAATAACGCTGGCGCGCTCTTAAACATGCAAGTCAACGCGAGTAG 102
Qy 61 caatcacctagcggcgaacgggtgaataacacgt -ggtaatcttctccagctctgggata 119
Db 103 CAATACTCAGCGCGCAACGGGTGAGTAAACGTGGGTAAATCTTCCTNCGAGTCTGGGATA 162
Qy 120 actttccgaagaagaactaaaccggtatgctctgttggatcacagaattttgatgata 179
Db 163 ACTTTCCGNAAGGGAAGCTAATCTGATGTCCTCCGAGAGATCAATAAGATTTTTCGGGTA 222
Qy 180 aagattttattgttgatagagcccgccgagattagctagttgtgaggttaattggctc 239
Db 223 AAGATTCATTGCTCGAGATGAGCCCGCTCCGATTAGTTCGTAAGGTAAAGGCTT 282
Qy 240 accaagcgacatcggttagccggcctgagaggggtgctcggccacaatggaactgagaca 299
Db 283 ACCAAGCGACATCGGTAGCGCGCTCAGAGGGGTGTTCCGGCCACAATGGAACCTGAGACA 342
Qy 300 cgggtccactactcctcgggagcgagcttaagaatcttctcaatggggggaacccctga 359
Db 343 CGGTCCATACT-CTACGGGAGGACAGCTTAAGATCTTGCTCAATGGGGGGAACCCCTGA 401
Qy 360 agcagcgacgcgcgcggaacgaagaagctcttcgagttgtaaaagttcattaggcca -ggaa 418
Db 402 AGCAGCGACGGCGGTGAACGATGAAGGCTCTCGGATTTGTAAGGTTCAATGAAGCAGGAA 461
Qy 419 aataagcagcaatgtgatgtgtaacctgctaaagcacgcgctaaactacgtgccagca 478
Db 462 AAATAAGCAGCAATGTGATGATGCTACCTTGCTTAAGCACCAGCTAAGTACGTGCCAGCA 521
Qy 479 gccgcggttaacgtatggtgcaagcgttgttcggaatcattggcggttaaaaggtgcgta 538
Db 522 GCCGCGGTAAATACGTATGTCGCAAGCGTGTTCGGAATCATTTGGCGGTAAAGGGTGCCTA 581
Qy 539 ggcggtattgaagtcaggtgtaaaactcggcgtcaacccgtgacctgcaactgaaac 598
Db 582 GCCGCACATGTAAGTCAGGTGTGAACCTGGGGGCTCAACTCCAGCGCTGCACCTGGAAC 641
Qy 599 tacaagctgagttggagaggaagcgaatgccaggtgtagcgtgaaatgcgtag 658
Db 642 TATCTGTCTGAGTTTGGGACAGCAAGTGAATTCACAGTGTAGCGGTGAATTCGGTAG 701
Qy 659 atatctggaggaacacacagtggtggaagcgactgctggtctcaaaactgacgtgagcca 718
Db 702 ATATCTGGAGGAACACCAGTGGCGGAAGCGACTTGTCTGGCTTAAACTGACGCTGAGGCA 761
Qy 719 cgaagcgtggtgtagtaaacgggtagataccctcggttaatccacgcctcaaacgtttgtc 778
Db 762 CGAAGCGTGGGTAGTGAACGGGATAGATNCCCGGTAAATCCACGCCCTTAACGTTGTC 821
Qy 779 taccagttgttgggggttttaacctcagtaacgaactaacggtatgaagtagaccgct 838
Db 822 TACCAGTTGTTGGGGGTTTTAACCTCAGTAACCAA-CTAACGGATTAAAGTAGACCGCT 880
Qy 839 gggcactatgctcgaagagtgaaactcaaaaggaattgacgggggtccgcacaaagcggtg 898
Db 881 GGGACATATGCTCGCAAGTGAAGTCAAGAGGAATTGACGGGGGTNNNCACAAAGCGGTG 940
Qy 899 gagcatgtggttgaattcgaatgatacccccacacccctcaactgggttgacatggatctg 958
Db 941 GAGCATGTGGTTTTAATTCGATGATACCGGAACAAACCTCACCTAGGCTTGACATGAGTGG 1000
Qy 959 aatcatgtagagatatagaccttcctcggcagatctcagagtgctgcatggtgtctca 1018
Db 1001 AATCATGTAGAGATACATAGAC-CTTCGGGGCGCTTTCACAGGTGCTGTCATGTTCTCGTCA 1059
Qy 1019 gctcgtgctgagatattgggttaagttcccgcaacgagcgaacccctatcgtatttg 1078
Db 1060 GCTCGTGTCTGATGATGTTGGGTTAAGTTCGGCAACGAGGCAACCCCTCACCTTATGTNG 1119
```

```
Qy 1079 cta-ccctaaattggcactggtacgaactgctggttcaaaaacgcgaggaagcgaggga 1137
Db 1120 CCATCATTCAGTTGGCACTCGTAAGGAACCTGCCGCTGCAAAACGGAGGAGCGGGGA 1179
Qy 1138 tgaagtcataatctcatggtccctttatgtccaggggccaacacgctgctacaatgcccagata 1197
Db 1180 TGAGTCAAAATCCTCATGGCCTTTATGTCTAGGGCAACACACAGCTGCTACANTGGCCGTA 1239
Qy 1198 cagagggctcgccaactcgcaagaggggagctaatctctctaaagtcgggtcccgattcggtatt 1257
Db 1240 CAAAGGGTAGCAACTCGCGAGGGGAGCTAATCTCAAAAAGCGGTGCCAGTTCGGATT 1299
Qy 1258 ggggtcttgcgaactgcaccccatgaagtcggaatcgctagtaataatcgagatcagcatgccg 1317
Db 1300 GGAGTCTCAACTCGACTCCATGAAGTCGGATCGTAGTAATCCGGATCAGCATGCCG 1359
Qy 1318 cgggtgaatacgttccccggacctgtacacacccgcgcgtcacaccacctgagtgggagca 1377
Db 1360 CGGTGAATACCTTCGGGACCTTGTACACACCGCGCTCACACCACCTGAGTGGGAGCA 1419
Qy 1378 cccgaagtggtcttgttaaccgtgaagagagagagagactactaagtgaaactcgttaagggg 1437
Db 1420 CCCGAAGTGGCTTTGGCCACCGCAAGGAGCAGACTACTAAGGTGAACCTCGTGAAGG 1479
Qy 1438 ggtgaagtcgtaacaaggtacc 1459
Db 1480 GGTGAAGTCGTAAACAAGGTAGC 1501
```

RESULT 10

LB16SRDNP

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

rRNA

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches 1343; Conservative

84.9%; Score 1254.6; DB 2; Length 1425;

Pred. NO. 2.5e-13;

Mismatches 76; Indels 4; Gaps 4;

Qy 16 ctaacgctggcgctgctttaaacaatgcaagtc-gagcgggtagcaatacctagcgc 74

Db 1 STARCGTGGCGCGCTCTTAAACATCCCAATCCAAAGCGGAGTAGCAATACTACGCGG 60

Qy 75 gaacgggtgagtaaacagct-ggtaattcttcccgagctcgggataacttccgaagga 133

Db 61 GAACGGGTGAGTAACAGCTGGGTAAATCTTCTCCGAGTCTGGGATAACTTTCGGAAGG 120

LB16SRDNP 1425 bp DNA BCT 02-DEC-1993
L.borgpetersenii gene for 16S ribosomal RNA (partial).
221630
221630.1 GI:433579
16S ribosomal RNA.
Leptospira borgpetersenii.
Bacteria; Spirochaetales; Leptospiraceae; Leptospira.

Hooley, J. V.
Phylogeny of Leptospiraceae and related Spirochaetes

Unpublished
2 (bases 1 to 1425)

Hooley, J. V.
Direct Submission

Submitted (09-FEB-1993); Hooley J.V., Public Health Laboratory

Service, Leptospira Reference Laboratory, Stonebow Road, Hereford,

Herefordshire, United Kingdom, HR1 2ER

Location/Qualifiers
1. 1425

/organism="Leptospira borgpetersenii"

/strain="JAVANICA VELDRAT BATAVIA 46"

/db_xref="taxon:174"

<1..>1425

/product="16S ribosomal RNA"

369 a 327 c 430 g 297 t 2 others

QY 134 aagctaataaccggatagctctgttgatcaacaagatttgataggttaagattatttgctt 193
DB 121 GAGCTTAATACTGGATAGTCCCGAGAGGCTATAGGATTTTTCGGGTAAAGATTATTGCTC 180
QY 194 ggaatgaaccccgccgagattagctagctgaggttaaggtcaccacagcgacgat 253
DB 181 GGAGATGAGCCCGCCCGGATTTAGCTAGTTGGTGAAGTAATGGCTTACCACAGCCACGAT 240
QY 254 cggtagcggcctgagaggggttcggccgacacaaatggaactgagacaggtccatactct 313
DB 241 CGGTAGCCGGCTGAGAGGGTCTTCGGCCACAAATGGAAGTGAACAGCGTCCATCTCT 300
QY 314 acggagggcagcagtagtaagaattctgctcaatgggggaacccctgaagcagcagcgcg 373
DB 301 ACGGAGGGCAGCAGTTAAGAAATCTTGCTCAATGGGGGAACCCCTGAAGCAGCAGCCGCG 360
QY 374 gtaacgaagaaggctcttcggattgaaagttcattagga-ggaaaaataagcagcaat 432
DB 361 GTGAACGATGAAGGTCTTCGGATTGTAAAGTTCAATAAGCAGCGGAATAAGCAGCGAT 420
QY 433 gtgatgtgtaacctgcctaaagcagcccggttaactacgtgccagcagcgcggtaacag 492
DB 421 GTGATGATGTTACCTGCCTAAAGCAGCCGCTAATCTAGCTGCCAGCAGCCGCGTAAACG 480
QY 493 tatgtgcaagcgttcttcggaaatcattggcgtaaggggtcgtaggcgatttgtaag 552
DB 481 TATGTTGAAGCGTTGTTGCGGAATCATTTGGCGTAAAGGGTGGTGAAGCGACATGTAG 540
QY 553 tcagctgtgaacactgcggctcaaccgtgacctgcactgaactgaactcaagctcggagt 612
DB 541 TCAGTGTGAACACTGCGGGCTCAACTCCAGCCCTGCACCTGAACATATGCTCTGGAGT 600
QY 613 ttggagagggcaagtggaaattccaggttagcgggtgaatgcgttagatctctggaggaa 672
DB 601 TTGGGAGAGGCAAGTGAATTCACGGTGTAGCGGTGAATGCTAGATATCTGGAGGAAC 660
QY 673 accagtggcgaaggcagcttgctgctcaaaactgacgtgaggcagcaaaagcgtgggta 732
DB 661 ACCAGTGGCGAAGGGCACTTGTGCGCTAAACTGACGCTGAGGACGGAACGGTGGGTA 720
QY 733 gtaacgggattagatacccccgttaacacccgttaacacgttgtctaccagttgttgg 792
DB 721 GTGAACGGGATTAGATACCCCGGTAAATCCACGCCCTAAACGTTGTCTACCACTTGTGG 780
QY 793 ggttttaaccctcagtaacaaacctaaaggatgaagttagaccgacctgggagctatgctg 852
DB 781 GGTTTTAAACCTCAGTACGAACCTAACGGATTAAGTAGACCGCCCTGGGGACTATGCTCG 840
QY 853 caagagtaaaactcaagaaattgacggggtccgcacaaagcggtagacatgtgttta 912
DB 841 CAAGAGTGAACCTCAAGGAATTGACGGGGTCCCGACAAAGCGGTGGAGCATGTGTTTA 900
QY 913 attcagatatacccaaaacctcacctgggcttgacatggatcgtgaatcatgtagagat 972
DB 901 ATTCGATGATACCGGAAACCTCACCTAGGCTTGACATGGAGTGAATCATGTAGAGAT 960
QY 973 atatgaccttcgggcagatcacagggtgctgcatgggttgctgacgtcgtgctgag 1032
DB 961 ACATGAGCCTTCGGGCGGCTTCACAGGTGCTGCATGTTGTGCTGAGCTCGTGTGTAG 1020
QY 1033 atgttgggttaagttccgcacagcagcgaacccctatcgtatgttgcta-ccctaaagttg 1091
DB 1021 ATGTTGGGTAAAGTCCCGCAACGAGCGCAACCTCACCTTATGTTGCCATCATCAGTTG 1080
QY 1092 ggcactggtacgaactgcgggtgacaaacccggaggagcgggatgacgtcgaactcct 1151
DB 1081 GGCACCTGTAAGAACTGCCGTGACAAACCGAGGAGGCGGGATGACGTCAAACTCT 1140
QY 1152 catggcctttatgtccagggccacacagctgctcaaatggccgatacagaggggtgcgcaa 1211
DB 1141 CATGGCCTTTATGCTAGGGCAACACACGCTGCTACAATGGCCGGTACAAAGGGTAGCCAA 1200
QY 1212 ctgcgaagaggagactaactctctaaagtcggtcccgagttcgggtcgcgaactc 1271

DB 1201 CTCGGCAGGGGAGGTAAATCTCAAAAGCCGGTCCAGTTCCGATTTGGAGTCTGCAACTC 1260
QY 1272 gacccatgaagtcggaatcgctagtaatcgcgatcagcatgcgagtcgagtcgaactgctc 1331
DB 1261 GACTCCATGAAGTCGGAATCGCTAGTAATTCGGGATCAGCATGCCGGGTGAATAGCTTC 1320
QY 1332 cggacctgtgtacacacgccccgtcacacacctgagtggggagcaccacgaagtgtctt 1391
DB 1321 CGGAGCTTGTACACACCGCGGTACACACCTAGTGGGAGCACCAGGAGTGTCTT 1380
QY 1392 tattaacctgaaggagacagactactaagtgtaactcgtaaagg 1436
DB 1381 TGCCAACCCCAAGGAGCAGACTACTAAGGTGAACCTCGTGAAGG 1425

RESULT 11
WB4616S
LOCUS WB4616S 1486 bp DNA BCT 18-FEB-1995
DEFINITION Leptospira weilii Sarmin 16S rRNA gene, partial sequence.
ACCESSION U12673
VERSION U12673.1 GI:558932
KEYWORDS
SOURCE Leptospira weilii.
ORGANISM Leptospira weilii.
Bacteria; Spirochaetales; Leptospiraceae; Leptospira.
REFERENCE 1 (bases 1 to 1486)
AUTHORS Ralph D. and McClelland, M.
TITLE Phylogenetic evidence for horizontal transfer of an intervening sequence between species in a spirochete genus
JOURNAL J. Bacteriol. 176 (19), 5982-5987 (1994)
MEDLINE 95014031
REFERENCE 2 (bases 1 to 1486)
AUTHORS McClelland, M.
TITLE Direct Submission
JOURNAL Submitted (25-JUL-1994) Michael McClelland, California Institute of Biological Research, 11099 North Torrey Pines Road, La Jolla, CA 92037, USA

FEATURES
source
1..1486
/organism="Leptospira weilii"
/strain="serogroup Icterohaemorrhagiae, subserogroup Sarmin, serovar sarmin, strain Sarmin"
/isolate="WB46"
/db_xref="taxon:28184"
<1..1486
/product="16S rRNA"
BASE COUNT 384 a 332 c 452 g 315 t 3 others
ORIGIN

Query Match 84.8%; Score 1251.8; DB 29; Length 1486;
Best Local Similarity 94.0%; Pred. No. 2.7e-13;
Matches 1375; Conservative 0; Mismatches 80; Indels 7; Gaps 7;

QY 1 gatcatgctcgaactaacgctgcggcgctcttaaacatgcgaagtcgagcgggtag 60
DB 14 GATCTGGCTCAGACTAACGCTGGCGGCGCTTTAAACATGCAAGTCAAGCGGATAG 73
QY 61 caatacctagcggcaacggtgagtaacacgt-ggtaactcttcctccagtcgggata 119
DB 74 CAATATCCACGGCGGAACGGGTGAGTAACACGTTGGGTAAATCTTCCCGAGTCTGGGATA 133
QY 120 acttcccgaaaggaagcctaaccgcatagctcgtttggatcacagaattgtaggta 179
DB 134 ACTTTCCGAAGGGAGAGCTTAATCTGGATAGTCCCGAGAGGTCATATGATTTTCGGGTA 193
QY 180 aagattattgtttgagatgagccgcggcgagattagctagttggtgaggttaagtcct 239
DB 194 AAGATTATTGTGCGAGCTGAGCGCGCGCGGATTAGCTAGTTGTTGAGTAATGGCTC 253
QY 240 accaaggcgacatcgttagccgctgagaggggtgtccggccacaaatgaactgagaca 299

Db 254 ACCAAGCGCAGCATCGGTAGCCGCTGAGAGGGTGTTCGGCCACAATGAACTGAGACA 313
Qy 300 cggctccatactcctacgggagcagcagtaagtaattctgctcaatggggaacccctga 359
Db 314 CGGTCCATACT-CTACGGGAGCGCAGCAGTAAAGAACTTTCCTCAATGGGGGAAACCCCTGA 372
Qy 360 agcagcagcgcgcgtgtaacgaagaaggtcttcggattgtaaaagtctcatagga-ggaa 418
Db 373 AGCAGCGACGCCGCTGAACCATCAAGGTCCTCGGATTTAAAGTCAATTAAGCAGGGAA 432
Qy 419 aataagcagcaatgtgatgtatgtactgcctaaagcacccggctactacgtgcacga 478
Db 433 AAATAAGCAGCAATGTGATGATGTACTGCTTAATCACCGGCTAACTAGTGCCAGCA 492
Qy 479 gcccggttaatacatgtgtcgaagcgtgttcggaatcattggcgtgaaggtgcga 538
Db 493 GCCCGGTGAATACGTATGCTGCAAGCGTGTTCGGAATCATTTGGCGGTAAAGGTGGCTA 552
Qy 539 ggcggatttgaagtcaggtgtgaaactcggcgtcaaccctggcctgcacttgaac 598
Db 553 GCGGCACATGTAAGTCAGTCTG-AACTGCGGGCTCAACTCGCAGCCTGCACTTGAAAC 611
Qy 599 tacaagtctgagttggagagcaagtggaattccaggtgtagcgtgaaatgcgtag 658
Db 612 TATGTCTGAGTTTGGGAGAGCAAGTGGAAATCCAGGTGTACGGGTGAATGCGGTAG 671
Qy 659 atactgagaacaccagtggaagcgacttgcgtgcacaaactgacgtgagga 718
Db 672 ATATCTGAGGAGAACACAGTCGCGAAGCGACTTGCCTGGCCCTAAACTGACGCTGAGCA 731
Qy 719 cgaagcctgggttagtaaacgggtagatataccocgggtaataccacgcccataacgtgtc 778
Db 732 CGAAACGCTGGGTACTGAACGGGATAGATACCCCGGTAAATCCACGCCCTAAAGCTTGTCT 791
Qy 779 taceagttgtgggggttttaacctcagtaacgaacacagcgaattagtagccgct 838
Db 792 TACCAGTTGTTGGGGGTTTTAACCTCTAGTACGAA-CTAACGGATTAAGTAGACCGCT 850
Qy 839 ggggactatgctcgaagtgaaactcaaggaattgacgggggtccgcacaagcggtg 898
Db 851 GCGGACTATGCTCGCAAGAGTCAAACTCAAGGAATTCACGGGGTNNNCAACGCGGTG 910
Qy 899 gagcatgtggttaattcgatgatacccccaaaacctcaacctggcgttgacatgacgtg 958
Db 911 GAGCATGTGGTTAATTCGATGATACGGCGAAACCTCACCTAGGCTTGACATGGAGTGG 970
Qy 959 aatcatgtagatatatagccttcggcagatccacaggtgctgcatggtgtgcgtca 1018
Db 971 AATTATGTAGATACATGAG-CTTCGGGCCCTTCACAGGTGCTGCATGTTGTCGTCA 1029
Qy 1019 gctcgtgtgtgagatgttgggttaagtcocgcgaacgagcgaacccctatogtatgtg 1078
Db 1030 GCTCGTGTCTGAGATGTGGGTTAAGTTCGCCCAACGAGCGCAACCCCTCACCTTATGTG 1089
Qy 1079 cta-cccttaagttggcactggtacgaactcgcggtgacaaacggaggaagcgagga 1137
Db 1090 CCATCATTTAGTTGGGCACTCGTAAAGAACTCCCGGTGACAAACCGGAGGCGGGGA 1149
Qy 1138 tgacgtcgaatcctcatgtgaccttattgtccagggccacacagctgctacaaatggccgata 1197
Db 1150 TGACGTCAANTCTTATGGCCCTTATGTCTAGGGCAACACACGCTGCTACATGGCCGTA 1209
Qy 1198 cagagggctgcgaactcgaagggagagtaattctctaaaaagtcggtccagttcggatt 1257
Db 1210 CAAAGGCTAGCCAACTCGCGAGGGGAGCTAATCTCAAAAGCCGCTCCAGTTCCGATT 1269
Qy 1258 ggggtctgaactcgaccccatgaagtcgaatcagtaatacgcgaatcagcatgcg 1317
Db 1270 GGAGTCTGCAACTCGACTCCATCAAGTCGGAATCGCTAGTAATCGCGGATCAGCATGCCG 1329
Qy 1318 cgggtgaatcgttcccgacctgttacacacccgcccgtcacacccactgagttggggagca 1377
Db 1330 CGGTGAATACGTTCCGGGACCTTGTACACACCGCCGCTCACACCACCTGAGTGGGGAGCA 1389

Qy 1378 ccgaagtgatcttgttaaccgttaaggacagactactactaaagtgaacctcgtaaaagg 1437
Db 1390 CCCGAAGTGTCTTTCACCAACCCCAAGGAGCAGACTACTAAGTGAACCTCGTGAAGG 1449
Qy 1438 ggtgaagtcgtaacaagggtacc 1459
Db 1450 GGTGAAGTCGTAACAAGGTAGC 1471

RESULT 12

LK16SRDNP LK16SRDNP 1419 bp DNA BCT 02-DEC-1993
LOCUS L.kirschneri gene for 16S ribosomal RNA.
DEFINITION
ACCESSION 221628
VERSION 221628.1 GI:433582
KEYWORDS 16S ribosomal RNA
SOURCE Leptospira kirschneri.
ORGANISM Bacteria; Spirochaetales; Leptospiraceae; Leptospira.
REFERENCE 1 (bases 1 to 1419)
AUTHORS Hooke, J. V.
TITLE Phylogeny of Leptospiraceae and related Spirochaetes
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1419)
AUTHORS Hooke, J. V.
TITLE Direct Submission
JOURNAL Submitted (09-FEB-1993) Hooke, J. V., Public Health Laboratory
Service, Leptospira Reference Laboratory, Stonebow Road, Hereford,
Herefordshire, United Kingdom, HR1 2ER

FEATURES

source
1. 1419
/organism="Leptospira kirschneri"
/strain="CYNOPTERI 3522 C"
/db_xref="taxon:29507"
<1..>1419

BASE COUNT 369 a 326 c 423 g 299 t 2 others
ORIGIN

Query Match 84.4%; Score 1246.6; DB 2; Length 1419;

Best Local Similarity 94.2%; Pred. No. 3.4e-13;
Matches 1336; Conservative 2; Mismatches 76; Indels 5; Gaps 4;

Qy 16 ctacagctggcgccgtctttaaactgc--aagtcagcggggtgacataccatagcgg 73

Db 1 STARCGTGGCGCGGCTTTAAACATTCCAAGTCCCAAGCGGAGTAGCAATACTACGCGG 60

Qy 74 cgaacgggtgagtaacacgt-ggtaactctcctccagctcggataacttccgaaagg 132

Db 61 CGAACGGTTGAGTACACGTGGGTAACTCTCTCCGAGTCTGGGATAACTTTCCGAAGG 120

Qy 133 aaagcctaataccgagtagctctgttgatcacacaagatttgataggtaaaagattattgct 192

Db 121 GAAGCTAATACTGGATGTGCCGAGAGATCATAGATTTTTCGGGTAAAGATTTATTGCT 180

Qy 193 tggagatgagcccgccggttagcttagctgttggtagagtaattggtcaccacgaagcaga 252

Db 181 CGGAGATGAGCGCGGCTCCGATTAGTATTGGTGGAGTAAGGCTCACCAGGCGCAGA 240

Qy 253 tcggttagccgacctgaagaggtgccggccacaatgggaactgaacacggtccatctacc 312

Db 241 TCGGTAGCGCGGCTTGAGAGGGGTTCGGCCACAAATGGAACTGAGACGCTGCATCTAC 300

Qy 313 tacggagcgacagcttaagaattctgctcaatggggaaacccctgaagcagcgacgcg 372

Db 301 TACGGGAGCGACAGCTTAAGAATCTTGCTCAATGGGGGAACCCCTGAAGCAGCGACGCG 360

Qy 373 cgtgaacgaagaaggtcttcggattgtaagtctattaggca-ggaaaaataagcagcaa 431

Db 361 COTGAACGATGAAGGTCTTCGGATTGTAAAGTTCTAGTAAGCGGGGAAAAATAAGCAGCA 420


```
Db 561 GCGGACATGTAACTCAGGTGTG-AAACTGCGGGCTCAACTCGCAGCCTGCACCTTGAAC 619
Qy 599 tacaagtctgagtttgggaagcaagtggaattccaggtgttagcggtagaatacgtag 658
Db 620 TATGTCTGTGAGTTTGGAGAGCAAGTGAATTCAGGTGTAGCGGTGAATTCGCTAG 679
Qy 659 atactgaggaacacacagtggaagcgacttgcctcacaactgacgtgagga 718
Db 680 ATATCTGGAGGAACACCACTGCGGAAGCGACTTGTCTGGCCCTAAACTGACGCTGAGGCA 739
Qy 719 cgaagcgtggttagtaaaacgggattagatacccccgggtaattccacgcccctaaacgtgtgc 778
Db 740 CGAAGCGTGGGTAGTGAACGGGATTAGATACCCCGGTAACTCCAGCCCTAAACGTTGTC 799
Qy 779 taccagttgttgggggttttaaccctcagtaacgaactaacgattaaagtagaccgct 838
Db 800 TACCAGTTGTTGGGGGTTTTAACCTCAGTAACGAA-CTAACGGATTAAAGTAGACCGCCT 858
Qy 839 gggactatgctcgaagagtgaaactcaaaaggaattgacgggggttcgcacaagcgggtg 898
Db 859 GGGGACTATGCTCGACAGAGTGAACCTCAAAAGGAATTGACGGGGGTNNNCACAAAGCGGTG 918
Qy 899 gagcatgtggttaattcgtatgatacccaaaaacccctcacctggccttgacatgattctg 958
Db 919 GAGCATGTGTTTAATTCGATGATACGCGAATAACCTCACCTAGCTTGACATGAGTGG 978
Qy 959 aatcatgtagagatatagacctctgggcagatctacagtgctgcatgattctgctca 1018
Db 979 AATCATGTAGAGATACATAGAG-CTTGGGGCGCTTCACAGGTGCTGATGTTGTGCTGCA 1037
Qy 1019 gctcgtgctgagatgttgggttaagtcgcgcgaactgacgaaccctatcgtattgtg 1078
Db 1038 GCTGCTGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCTTCACCTTATGTTG 1097
Qy 1079 cta-cccttaagttgggcactggtacgaactgcggtgacaaaacggaggaagcgggga 1137
Db 1098 CCATCATTTTACTTGGGCACTCGTAAGGAACCTCCCGGTGACAAACCGGAGGAGCGGGGA 1157
Qy 1138 tgacgtcgaactcctcatggtccttatgttccagggccacacacgtgctacataggccgata 1197
Db 1158 TGAGCTNNNNNNNNNNCCCTTTATGTCTATGGCAACACACACGCTGCTACAATGGCCGGTA 1217
Qy 1198 caggggtgcgaactcgaagagggagctaatctctaaaagtcggtcccgattcggatt 1257
Db 1218 CAAGGTGAGCAACTCCGAGGGGGAGCTAATCTCAAAAAGCCGCTCCCAAGTTCGGATT 1277
Qy 1258 ggggtctgcaactgcaccccatgaagtgcggaatcgctagtaatacgcggatcagcatgccg 1317
Db 1278 GGAGTCTCAACTCGACTCCATCAAGTTCGGAATCGTAGTAATCGCGGATCAGCATGCCG 1337
Qy 1318 cgggtgaatacgttcccgacattgtacacaccccgctcacacacactgagtgaggagca 1377
Db 1338 CGGTGAATACGTTCCCGACCTTTACACACCCCGCTACACACCCCTGAGTGGGGAGCA 1397
Qy 1378 cccgaagtggcttcttgaacgcgtaagagagacagactactaagtgaaactcgtaaaagg 1437
Db 1398 CCCGAGTGGTCTTTGGCAACCGTAAGGAAGACAGACTACTAAGGTGAACCTCGTGAAGG 1457
Qy 1438 ggtgaagtcgtacaaggtacc 1459
Db 1458 GGTGAAGTCTGTAACAAGGTAGC 1479
```

RESULT 14

```
LM16SRDX
LOCUS LM16SRDX 1418 bp DNA BCT 02-DEC-1993
DEFINITION L.meyeri gene for 16S ribosomal RNA (partial).
ACCESSION Z21648
VERSION Z21648.1 GI:433585
KEYWORDS 16S ribosomal RNA.
SOURCE Leptospira meyeri.
ORGANISM Bacteria; Spirochaetales; Leptospiraceae; Leptospira.
```

```
REFERENCE 1 (bases 1 to 1418)
AUTHORS Hookey,J.V.
TITLE Phylogeny of Leptospiraceae and related Spirochaetes
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1418)
AUTHORS Hookey,J.V.
JOURNAL Direct Submission
FEATURES
source 1..1418
/organism="Leptospira meyeri"
/strain="RANARUM ICE"
/rna /db_xref="taxon:29508"
<1..>1418
/product="16S ribosomal RNA"
BASE COUNT 372 a 324 c 421 g 299 t 2 others
ORIGIN
```

```
Query Match 83.8%; Score 1238.2; DB 2; Length 1418;
Best Local Similarity 94.2%; Pred. No. 4.6e-13;
Matches 1337; Conservative 2; Mismatches 75; Indels 5; Gaps 5;
```

```
Qy 16 ctaacgctggcggcgctcttaaacatgcaagtcgagcggggtagcaatacctagcggc 75
Db 1 STARCGTGGCGGGCGCTTTAAACATCCAA-TCAAGCGGAGTAGCAATACTACAGCGCG 59
Qy 76 aacgggtgagtaaacagt-ggtaatcttccctccagtcctggataaacttccgaagaa 134
Db 60 AACGGGTGAGTAACACGCTGGGTATCTCCCGAGTCTGGGATAACCTTTTCGAAGGGA 119
Qy 135 agctaataccggatgctctgttgatcacagaatttgataggttaagaattattgcttg 194
Db 120 AGCTAATACTAGTAGTCCCGAGAGATCATAAAGATTTTTCGGGTAAAGATTCATTGCTTG 179
Qy 195 gagatgagccggcgccgattagctagttgttgaggttaattggctcaccagggcagcgc 254
Db 180 GAGATGAGCCCGCTCCGATTAGCTAGTTGGTGGATTGCTCACCAGCGGACGATC 239
Qy 255 ggtagcggcctgaggggtgctccggccacaatggaactgagacacgctccatacctcta 314
Db 240 GGTAGCCGCTGAGAGGCTGTTCCGCCCAATGGAACCTGAGACACGCTCATACTCTTA 299
Qy 315 cgggagggcagcagttagaattcttctcaatgggggaacccctgaagcagcagcgcgcg 374
Db 300 CGGGAGGCAGCAGTTAACAATCTTGTCAATGGGGGAACCTTGAAGCAGCGACGCCGCG 359
Qy 375 tgaacgaagaaggtcttcggattgtaaagttcattagga-ggaaaaataagcagcaatg 433
Db 360 TGAACGATGAAGGCTTTCGGATTGTAAAGTTCAATAAGCAGGGAATAAGCAATG 419
Qy 434 tgatgattgtaccttcctaaagcaccggctactacgtgcgcagcagccggttaatacgt 493
Db 420 TGATGATGTTACCTCCCTAAAGCACCAGCTAACTACGTGCCAGCAGCCGCTAATAGT 479
Qy 494 atggtgaagcgttcttcgggaatcattggcgtaaaaggtgctagcggtattgtaagt 553
Db 480 ATGTTGAAGCGTTTTCGGAAATCATTTGGCGTAAAGGTTGCTAGCGGACATGTAAGT 539
Qy 554 caggtgtaaaactcgggctcaaccggtgctcgaactgaaactacaagtctgaatt 613
Db 540 CAGGTGTGAACACTGGGGGCTCAACTCCAGCCTTCGACTTGAACTATGTGCTCGAGTT 599
Qy 614 tgggagggcgaagtgggaattccaggtgtagcgggtgaaatcgtagatatctggaggaaaca 673
Db 600 TGGGAGAGCGCAAGTGAATTCAGGTGTAGCGGTGAAATGCGTAGATATCTGGAGGAACA 659
Qy 674 ccagtgccgaaggcagcttctgctcaaaactgacgcgtgagggcagcaagcgtgggttag 733
Db 660 CCAGTGGCGAAGGCGACTTGTCTGGCCTAAAACTGACGCTAGGCGACGAAAGCGTGGGTAG 719
```

```

QY 734 taaacggattagataccccggtaataccacgcccctaaacgttggtgtgtatccagttgtgtggg 793
Db 720 TGAACGGGATTAGATACCCCGGTAAATCCACGCCCTAAACGGTTGTCTACAGTTGTGGG 779
QY 794 gtttaacccctcagtaacgaacctaacgattaaagtagacgcgcctgggactatcgc 853
Db 780 GTTTTAACCCCTCAGTAACGAACCTAACCGATTAACTAGACGCCCTGGGACTATGCTCGC 839
QY 854 aagadgaacactcaagaattagcgggggtccgcacaaagcgggtggagcatgtgtttaa 913
Db 840 AAGAGTGAACCTCAAGGAATTGACGGGGTCCGCACAAAGCGGTGGAGCATGTGTTTAA 899
QY 914 ttcgatgatacccaaaaacacctcactggcttgcacatggtatcgaatcatgtaagata 973
Db 900 TTCGATGATACCGGAACCTCACCTAGGCTTGACATGGAGTGAATCATGTAGAGATA 959
QY 974 tatgaccttggggcagattcacagtgctgcattgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1033
Db 960 CATGAGCCCTTCGGGCGGCTTCACAGTGTGTCATGGTTGTAAATCAGCTCGTCTGAGA 1019
QY 1034 ttttgggttaagtccgcgaacgagcgaacccctatctgtgtgtacc-ttaaagttgg 1092
Db 1020 TGTGGGTAAAGTCCCGCAACGAGCGCAACCTCACCTTATGTTGCCATCATTCAGTTGG 1079
QY 1093 gcaactgtacgaactccggtgacaaacccgaggaagcgggtgatgacgtcaaatctc 1152
Db 1080 GCATCTGTAAGAACTGCGCGGTGACAAACCGAGGAAGCGGGGTGATGACGTCAAATCCTC 1139
QY 1153 atggcctttatgtccaggccacacacgtgtacaaatggccgatacagaggtgcgaac 1212
Db 1140 ATGGCCTTTATGCTPAGGCAACACACGTGTACAAATGGCCGGTACAAAGGTAGCCCAAC 1199
QY 1213 tcgaaggaggagctaaatctctaaagtcggtccagttcggattggggtctgcaactcg 1272
Db 1200 TCGCAGAGGGGAGCTTAATCTCAAAACCCGGTCCCGAGTTCGGATTGGAGTGTGCAACTCG 1259
QY 1273 acccctgaagtcggaatcgctagtaatcgcgatccagatccagtcgcgcgtgaatcgttcc 1332
Db 1260 ACTCATGAGTCGGAATCGGTAGTAATCGGGATCAGCATGCCCGGTGAATAGTTCC 1319
QY 1333 cggacctgtcacaccccgctcacacacctgagtgaggagcaccggaagtgtcttt 1392
Db 1320 CGGACCTTGTACACACC-CCCGTACACACCTGAGTGGGAGCACCCGAAAGTGTCTTT 1378
QY 1393 gttaacctgaaggagacagactactaaagtgaaactcgt 1431
Db 1379 GCCAACCGCAAGGACGACTACTAAGGTGAACCTCGT 1417

RESULT 15
L116SRRN L116SRRN 1396 bp mRNA BCT 24-AUG-1993
LOCUS L.interrogans 16S ribosomal RNA.
ACCESSION 212817
VERSION 212817.1 GI:44003
KEYWORDS 16S ribosomal RNA; ribosomal RNA.
SOURCE Leptospira interrogans.
ORGANISM Leptospira interrogans.
REFERENCE 1 (bases 1 to 1396)
AUTHORS Bryden, J., Gatehouse, L., Gatehouse, J.A. and Hooley, J.V.
TITLE Automated PCR cycle sequencing of 16S ribosomal RNA genes
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1396)
AUTHORS Hooley, J.V.
TITLE Direct Submission
JOURNAL Submitted (19-JUN-1992) Hooley J. V., Public Health Laboratory
Service, Leptospira Reference Laboratory, Stonebow Road, Hereford,
Herefordshire, United Kingdom, HR1 2ER
LOCATION/Qualifiers
FEATURES
source 1. .1396
/organism="Leptospira interrogans"
/strain="icterohaemorrhagiae RGA"

```

```

/db_xref="taxon:173"
1. .1396
/standard_name="16S ribosomal RNA"
BASE COUNT 365 a 319 c 418 g 294 t
ORIGIN
Query Match 83.5%; Score 1232.8; DB 2; Length 1396;
Best Local Similarity 94.6%; Pred. NO. 5.6e-13;
Matches 1309; Conservative 0; Mismatches 72; Indels 3; Gaps 3;
QY 49 cgagcgggtgagcaataacctagcggcgaacgggtgagtaaacagt-ggtaattctctctcc 107
Db 13 CAAGCGGAGTAGCAATACTCAGCGCGCAACGGGTGAGTAACACGTGGGTAACTTTCCTCT 72
QY 108 gactctgggataaactttccgaaaggaagactaaataccggatagctctgttgatcacaaag 167
Db 73 GAGTCTGGGATAACTTTCGGAAGGGAAGCTTAATCTGGATGGTCCCGAGAGATCATAG 132
QY 168 attgataggtaaaattattgcttgagatgagcccgccgagcttagctagttgggtg 227
Db 133 ATTTTTCGGTAAAGATTATTGCTCGGAGATGAGCCGCGTCCGATTAGCTAGTTGGTG 192
QY 228 aggtaatggctcaccaaggcagatcgtgtagcggcgcctgagaggggttccggccacaat 287
Db 193 .AGGTAAGGCTCACCAAGCGCAGATCGGTAGCCCGCCTGAGAGGGTGTTCGGCCACAAT 252
QY 288 ggaactgagacacggtccatactctacggagggcagcaggttaagaatcttgctcaatgg 347
Db 253 GGAAGTGAAGACGCTTCTACTCTACGGGAGGAGCAGCATTAAGAATCTTGCTCAATGG 312
QY 348 gggaaacctgaagcagcgcgcgtgaacgaagaaggtcttcggattgtaaagtcca 407
Db 313 GGGGAACCTGAAGCAGGAGCGCCGCGTGACGATGAAGGTCTTCGGATTGTAAAGTTCA 372
QY 408 ttaggca-ggaaaaataagcagcaatgtagtgcctacccctcctaaagcagcgctaac 466
Db 373 GTAACAGGGAAATAAGCAGCAATGTATGATGGTACCTGCCCTAAAGCACCGGCTAAC 432
QY 467 tacgtgcagcagcgcggttaatacgtatggtgcaagcgttcttcggatcatcttggggt 526
Db 433 TACGTGCCAGCAGCGCGTAAATACGTATGTTGCAAGCGTTCGTAATCATTTGGCGGT 492
QY 527 aaagggtcgttaggcggtattgtaagtcaggtgtgaacactcgcggcctcaaccccgcc 586
Db 493 AAAGGTCGTAGCGCGACATGTAAGTCAGGTGTGAAACCTCGCGGCTCAACTCGCAGCC 552
QY 587 tgcactgaaactacaagctcggagtttggagagcgaagtgaatccagggttagcgg 646
Db 553 TGCACCTTGAACATATGTCTGGAGTTCGGGAGAGCGAAGTGAATTCAGGTGTAGCGG 612
QY 647 tgaatgcgtagatatactggaggaacacacagtcggcgaaggcgacttgcgtcctcaaac 706
Db 613 TGAATGCGTAGATATCTGGAGGAACACACAGTCGGCGAAGGCGACTTGTGGCTAAAAC 672
QY 707 gacgtgaggcagcaagcgttgggttagtaaacgggttagatatacccccggtaacccagcc 766
Db 673 GACGCTGAGGACGAAAGCGTGGGTAGTGAACGGGATTAGATACCCCGGTAAATCCACGCC 732
QY 767 ctaaaactgtctaccagtttgggggttttaaccctcagtaacgaacccaacccaagatta 826
Db 733 CTAACAGTTGTCTACCACTGTTCGGGGTTCCTTAAACCCCTCAGTAACCAACCTAACCGATTA 792
QY 827 agtagaccgctgggagactatgctcgaagatgaaactcaaggaattgcaggggggtcc 886
Db 793 AGTAGACCGCCTGGGACTATGCTCCGAAGAGTGAAGAACTCAAGGAATTCACGGGGGTC 852
QY 887 gcaaacggttgagcagctgtgtgttaattcgtatgatacccccaaacctcactgggctt 946
Db 853 GCACAAGCGGTGGAGCATGTGTTTAAATTCGATGATACGCGAAACCTCACCTAGGCTT 912
QY 947 gacatgactcgaatcatcgtagagatatatagccttcgggcagattcacagggtctgca 1006

```

```
Db 913 GACATGGAGTCGAATCATGTAGATACATGACGCTTCGGGCCGCTTCACAGGTGCTGCA 972
Qy 1007 tggttgtcgtcagctcgtcgtgagatgttggttaagtcgccgcaacgagcgcaacccc 1066
Db 973 TGGTTGTGCTCAGCTCGTCTCGTCAGATGTTGGGTTAAGTCCCGCAACGACGCAACCCCT 1032
Qy 1067 ktcgltatgttgcta-ccttaagtgggcactggtaagaaactgccggtgacaaacccgga 1125
Db 1033 CACCTTATGTTGCCATCATTCAGTTGGGCACCTCGTAAGGAACTGCCGGTGACAAACCGGA 1092
Qy 1126 ggaaggggggagatgacgtcaaatcctcatgaccttttatgtccaggggccacacagtacta 1185
Db 1093 GGAAGGGGGGATGACGTCAATCCTCATGGCCCTTTATGCTAGGGCAACACACGTGCTA 1152
Qy 1186 caatggccgatatcacagaggggtcgccaaactcgcaagaggagctaatctctaaagtgcgtc 1245
Db 1153 CAATGGCCGGTACAAAGGGTAGCCAACTCGGGAGGGGAGCTAATCTCAAAAATCCGGTC 1212
Qy 1246 ccagttcggattggggtctgcaactcgacccccatgaagtcgggaatcgctagttaatcgcg 1305
Db 1213 CCAAGTCGGATTGGAGTGTGCAACTCGACTCCATGCAAGTGGGAATCGTATGTAATCGCGG 1272
Qy 1306 atcagcatgccgcggtgtaatacgttcccgaccttgtacacacgcgcgtcacaccacct 1365
Db 1273 ATCAGCATGCCGCGGTGANTACGTTCCGGACCTTGTACACACCGCCGCTCACACCACCT 1332
Qy 1366 gagtggggagcaccgcgaagtgttcttttaaccgtaagagacagactactaaggtgaa 1425
Db 1333 GAGTCGGGAGCAGCACCAGTGGTCTTTGCCAAACCGCAAGGAGGAGACTACTAAGGTGAA 1392
Qy 1426 actc 1429
Db 1393 ACTC 1396
```

Search completed: March 27, 2001, 08:16:52
Job time: 5889 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 27, 2001, 08:17:03 ; Search time 2286.42 Seconds
(without alignments)
15.668 Million cell updates/sec

Title: US-09-380-826A-4
Perfect score: 7
Sequence: 1 tgttga 7

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1118133 seqs, 2558875100 residues

Total number of hits satisfying chosen parameters: 2236266

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:

- 1: gb_ba1.*
- 2: gb_ba2.*
- 3: gb_om.*
- 4: gb_ov.*
- 5: gb_ph.*
- 6: gb_pl1.*
- 7: gb_pl2.*
- 8: gb_pr1.*
- 9: gb_pr2.*
- 10: gb_pr3.*
- 11: gb_ro.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: em_fun.*
- 15: em_hum1.*
- 16: em_hum2.*
- 17: em_in.*
- 18: em_om.*
- 19: em_or.*
- 20: em_ov.*
- 21: em_pat.*
- 22: em_ph.*
- 23: em_pl.*
- 24: em_ro.*
- 25: em_sts.*
- 26: em_sy.*
- 27: em_un.*
- 28: em_v1.*
- 29: gb_ba3.*
- 30: gb_in1.*
- 31: gb_in2.*
- 32: gb_in3.*
- 33: gb_pl3.*
- 34: gb_pr4.*
- 35: em_ba1.*
- 36: em_ba2.*
- 37: em_btgl.*
- 38: em_btgt.*
- 39: em_btgs.*
- 40: em_btga.*
- 41: em_btgs.*
- 42: em_btgs.*
- 43: em_btgs.*

- 44: em_btgs.*
- 45: em_btgs.*
- 46: em_btgs.*
- 47: em_hum3.*
- 48: em_hum4.*
- 49: em_hum5.*
- 50: em_hum6.*
- 51: gb_pr5.*
- 52: gb_pr6.*
- 53: gb_pr7.*
- 54: gb_btgl.*
- 55: gb_btgs.*
- 56: gb_btgs.*
- 57: gb_btgs.*
- 58: gb_btgs.*
- 59: gb_btgs.*
- 60: gb_btgs.*
- 61: gb_btgs.*
- 62: gb_btgs.*
- 63: gb_btgs.*
- 64: gb_btgs.*
- 65: gb_btgs.*
- 66: gb_btgs.*
- 67: gb_btgs.*
- 68: gb_btgs.*
- 69: gb_btgs.*
- 70: gb_btgs.*
- 71: gb_btgs.*
- 72: gb_btgs.*
- 73: gb_btgs.*
- 74: gb_btgs.*
- 75: gb_btgs.*
- 76: gb_btgs.*
- 77: gb_btgs.*
- 78: gb_btgs.*
- 79: gb_btgs.*
- 80: gb_btgs.*
- 81: gb_btgs.*
- 82: gb_btgs.*
- 83: em_btgs.*
- 84: gb_btgs.*
- 85: gb_btgs.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	7	100.0	10	81	AR062865
C 2	7	100.0	13	81	A00432
C 3	7	100.0	14	51	AX009010
C 4	7	100.0	14	79	AX007733
C 5	7	100.0	14	79	AX007734
C 6	7	100.0	14	81	A25812
C 7	7	100.0	14	81	A40518
C 8	7	100.0	14	81	A88271
C 9	7	100.0	14	81	A88273
C 10	7	100.0	14	81	A89045
C 11	7	100.0	14	81	A90238
C 12	7	100.0	14	81	A90240
C 13	7	100.0	14	81	I26234
C 14	7	100.0	15	79	AX007732
C 15	7	100.0	15	81	A88272
C 16	7	100.0	15	81	A88466
C 17	7	100.0	15	81	A90239
C 18	7	100.0	15	81	A90433
C 19	7	100.0	15	81	AR033435
C 20	7	100.0	15	81	AR033480
C 21	7	100.0	15	81	I57664

c 22 7 100.0 15 81 I57709 Sequence 24
 c 23 7 100.0 16 81 A96937 Sequence 15
 c 24 7 100.0 16 81 A97828 Sequence 10
 c 25 7 100.0 16 81 A97828 Sequence 10
 c 26 7 100.0 16 81 AR084439 Sequence
 c 27 7 100.0 16 81 AR093881 Sequence 32
 c 28 7 100.0 16 81 I34946 Sequence 15
 c 29 7 100.0 16 81 I52073 Sequence 43
 c 30 7 100.0 17 81 A65762 Sequence 18
 c 31 7 100.0 17 81 A97904 Sequence 18
 c 32 7 100.0 17 81 AR046017 Sequence
 c 33 7 100.0 17 81 AR053055 Sequence
 c 34 7 100.0 17 81 AR065016 Sequence
 c 35 7 100.0 17 81 I04576 Sequence 6
 c 36 7 100.0 17 81 I13562 Sequence 4
 c 37 7 100.0 17 81 I15174 Sequence 23
 c 38 7 100.0 17 81 I32561 Sequence 25
 c 39 7 100.0 17 81 I37420 Sequence 43
 c 40 7 100.0 17 81 I50672 Sequence 4
 c 41 7 100.0 17 81 I53069 Sequence 81
 c 42 7 100.0 17 82 I89345 Sequence 2
 c 43 7 100.0 17 82 I94270 Sequence 43
 c 44 7 100.0 18 81 A07301 Nucleotide
 c 45 7 100.0 18 81 A07302 Nucleotide
 A45212 Sequence 89

ALIGNMENTS

RESULT 1
 AR062865/c AR062865 10 bp DNA PAT 29-SEP-1999
 LOCUS Sequence 13 from patent US 5843767.
 DEFINITION AR062865
 ACCESSION AR062865
 VERSION AR062865.1 GI:5990556
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 10)
 AUTHORS Beattie, K. L.
 TITLE Microfabricated, flowthrough porous apparatus for discrete
 detection of binding reactions
 JOURNAL Patent: US 5843767-A 13 01-DEC-1998;
 FEATURES Location/Qualifiers
 1..10
 /organism="unknown"
 BASE COUNT 4 a 4 c 0 g 2 t
 ORIGIN
 Query Match 100.0%; Score 7; DB 81; Length 10;
 Best Local Similarity 100.0%; Pred. No. 5.3e+05;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 t g t t g g a 7
 Db 7 T G T T G G A 1
 RESULT 2
 A00432/c A00432 13 bp DNA PAT 11-FEB-1993
 LOCUS Nucleotide sequence 7 from patent number WO9010459.
 DEFINITION A00432
 ACCESSION A00432
 VERSION A00432.1 GI:14504
 KEYWORDS
 SOURCE unidentified.
 ORGANISM unidentified.
 REFERENCE 1 (bases 1 to 13)
 AUTHORS
 TITLE PHARMACEUTICAL COMPOSITION USEFUL IN THE PREVENTION OR TREATMENT OF

JOURNAL PAPILLOMAVIRUS-INDUCED TUMOURS
 Patent: WO 9010459-A 7 20-SEP-1990;
 Transgene S.A.
 FEATURES Location/Qualifiers
 1..13
 /organism="unidentified"
 /db_xref="taxon:32844" 2 t
 BASE COUNT 4 a 3 c 4 g
 ORIGIN

Query Match 100.0%; Score 7; DB 81; Length 13;
 Best Local Similarity 100.0%; Pred. No. 5.1e-05;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 t g t t g g a 7
 Db 10 T G T T G G A 4

RESULT 3
 AX009010 14 bp DNA PRI 06-SEP-2000
 LOCUS Sequence 43 from Patent WO9963975.
 DEFINITION AX009010
 ACCESSION AX009010
 VERSION AX009010.1 GI:9996384
 KEYWORDS
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 14)
 AUTHORS Brysch, W., Schlingensiepen, K. H. and Schlingensiepen, R.
 TITLE A method for stimulating the immune system
 JOURNAL Patent: WO 9963975-A 16-DEC-1999;
 BIOLOGICAL RESOURCES (DE); BRYSCH WOLFGANG (DE); SCHLINGENSIEPEN KARL
 HERMANN (DE); SCHLINGENSIEPEN REIMAR (DE)

FEATURES Location/Qualifiers
 1..14
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 BASE COUNT 3 a 3 c 4 g 4 t
 ORIGIN

Query Match 100.0%; Score 7; DB 51; Length 14;
 Best Local Similarity 100.0%; Pred. No. 5e+05;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 t g t t g g a 7
 Db 3 T G T T G G A 9

RESULT 4
 AX007733/c AX007733 14 bp DNA VRL 06-SEP-2000
 LOCUS Sequence 275 from Patent WO9967428.
 DEFINITION AX007733
 ACCESSION AX007733
 VERSION AX007733.1 GI:9995430
 KEYWORDS
 SOURCE Aids-associated retrovirus.

ORGANISM Aids-associated retrovirus
 Viruses; Retroviridae; Retroviridae.
 REFERENCE 1 (bases 1 to 14)
 AUTHORS Stuyver, L.
 TITLE Method for detection of drug-selected mutations in the hiv protease
 gene
 JOURNAL Patent: WO 9967428-A 29-DEC-1999;
 INNOCENTICS NV (BE); STUYVER LIEVEN (BE)
 FEATURES Location/Qualifiers
 1..14
 /organism="Aids-associated retrovirus"


```
BASE COUNT      5 a      6 c      0 g      3 t
ORIGIN
|||||
/db_xref="taxon:11966"

Query Match      100.0%; Score 7; DB 79; Length 14;
Best Local Similarity 100.0%; Pred. No. 5e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgttggg 7
Db 12 TGTGGA 6

RESULT 5
LOCUS AX007734/c 14 bp DNA VRL 06-SEP-2000
DEFINITION Sequence 276 from Patent WO9967428.
ACCESSION AX007734
VERSION AX007734.1 GI:9995431
KEYWORDS Aids-associated retrovirus.
ORGANISM Aids-associated retrovirus
SOURCE Viruses; Retroid viruses; Retroviridae.
REFERENCE 1 (bases 1 to 14)
AUTHORS Stuyver, L.
TITLE Method for detection of drug-selected mutations in the hiv protease gene.
JOURNAL Patent: WO 9967428-A 29-DEC-1999;
INNOGENETICS NV (BE); STUYVER LIEVEN (BE)
FEATURES Location/Qualifiers
source 1..14
/organism="Aids-associated retrovirus"
/db_xref="taxon:11966"
BASE COUNT      5 a      6 c      0 g      3 t
ORIGIN
|||||

Query Match      100.0%; Score 7; DB 79; Length 14;
Best Local Similarity 100.0%; Pred. No. 5e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgttggg 7
Db 13 TGTGGA 7

RESULT 6
LOCUS A25812 14 bp DNA PAT 14-MAR-1995
DEFINITION polynucleotide 14C12.
ACCESSION A25812
VERSION A25812.1 GI:904780
KEYWORDS synthetic construct.
SOURCE synthetic construct.
ORGANISM artificial sequence.
REFERENCE 1 (bases 1 to 14)
AUTHORS Patent: FR 2680520-A 19 26-FEB-1993;
JOURNAL Location/Qualifiers
FEATURES source 1..14
/organism="synthetic construct"
/db_xref="taxon:32630"
BASE COUNT      4 a      4 c      2 g      4 t
ORIGIN
|||||

Query Match      100.0%; Score 7; DB 81; Length 14;
Best Local Similarity 100.0%; Pred. No. 5e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgttggg 7
```

```
Db 12 TGTGGA 6
|||||

RESULT 7
LOCUS A40518 14 bp DNA PAT 05-MAR-1997
DEFINITION Sequence 55 from Patent WO9425578.
ACCESSION A40518
VERSION A40518.1 GI:2296553
KEYWORDS unidentified.
SOURCE unidentified.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 14)
AUTHORS
TITLE ANTISENSE-OLIGONUCLEOTIDES FOR THE TREATMENT OF IMMUNOSUPPRESSIVE
EFFECTS OF TRANSFORMING GROWTH FACTOR--g(b) (TGF--g(b))
JOURNAL Patent: WO 9425578-A 55 10-NOV-1994;
BIOGNOSTIK GES (DE)
FEATURES Location/Qualifiers
source 1..14
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT      3 a      3 c      4 g      4 t
ORIGIN
|||||

Query Match      100.0%; Score 7; DB 81; Length 14;
Best Local Similarity 100.0%; Pred. No. 5e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgttggg 7
Db 3 TGTGGA 9

RESULT 8
LOCUS A88271 14 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 419 from Patent WO9833904.
ACCESSION A88271
VERSION A88271.1 GI:6736841
KEYWORDS unidentified.
SOURCE unidentified.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 14)
AUTHORS Brysch, W. and Schlingensiepen, K.
TITLE AN ANTISENSE OLIGONUCLEOTIDE PREPARATION METHOD
JOURNAL Patent: WO 9833904-A 06-AUG-1998;
BIOGNOSTIK GES (DE); BRYSCH WOLFGANG (DE)
FEATURES Location/Qualifiers
source 1..14
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT      5 a      4 c      0 g      5 t
ORIGIN
|||||

Query Match      100.0%; Score 7; DB 81; Length 14;
Best Local Similarity 100.0%; Pred. No. 5e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgttggg 7
Db 11 TGTGGA 5

RESULT 9
LOCUS A88273/c 14 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 421 from Patent WO9833904.
```

ACCESSION A88273
VERSION A88273.1 GI:6736843
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 14)
AUTHORS Brysch, W. and Schlingensiepen, K.
TITLE AN ANTISENSE OLIGONUCLEOTIDE PREPARATION METHOD
JOURNAL Patent: WO 9833904-A 06-AUG-1998;
BIOGHOSTIK GES (DE); BRYSCH WOLFGANG (DE)
FEATURES
source 1..14 Location/Qualifiers
BASE COUNT 6 a 4 c 0 g 4 t
ORIGIN

Query Match 100.0%; Score 7; DB 81; Length 14;
Best Local Similarity 100.0%; Pred. No. 5e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgttggg 7
Db 12 TGTGGA 6

RESULT 10
LOCUS A89045 14 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 1193 from Patent WO9833904.
ACCESSION A89045
VERSION A89045.1 GI:6737615
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 14)
AUTHORS Brysch, W. and Schlingensiepen, K.
TITLE AN ANTISENSE OLIGONUCLEOTIDE PREPARATION METHOD
JOURNAL Patent: WO 9833904-A 06-AUG-1998;
BIOGHOSTIK GES (DE); BRYSCH WOLFGANG (DE)
FEATURES
source 1..14 Location/Qualifiers
BASE COUNT 3 a 3 c 4 g 4 t
ORIGIN

Query Match 100.0%; Score 7; DB 81; Length 14;
Best Local Similarity 100.0%; Pred. No. 5e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgttggg 7
Db 3 TGTGGA 9

RESULT 11
LOCUS A90238 14 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 419 from Patent EP0856579.
ACCESSION A90238
VERSION A90238.1 GI:6738752
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 14)
AUTHORS Brysch, W. D. and Schlingensiepen, K. D.
TITLE An antisense oligonucleotide preparation method

JOURNAL Patent: EP 0856579-A 05-AUG-1998;
BIOGHOSTIK GES (DE)
FEATURES
source 1..14 Location/Qualifiers
BASE COUNT 5 a 4 c 0 g 5 t
ORIGIN

Query Match 100.0%; Score 7; DB 81; Length 14;
Best Local Similarity 100.0%; Pred. No. 5e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgttggg 7
Db 11 TGTGGA 5

RESULT 12
LOCUS A90240 14 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 421 from Patent EP0856579.
ACCESSION A90240
VERSION A90240.1 GI:6738754
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 14)
AUTHORS Brysch, W. D. and Schlingensiepen, K. D.
TITLE An antisense oligonucleotide preparation method
JOURNAL Patent: EP 0856579-A 05-AUG-1998;
BIOGHOSTIK GES (DE)
FEATURES
source 1..14 Location/Qualifiers
BASE COUNT 6 a 4 c 0 g 4 t
ORIGIN

Query Match 100.0%; Score 7; DB 81; Length 14;
Best Local Similarity 100.0%; Pred. No. 5e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgttggg 7
Db 12 TGTGGA 6

RESULT 13
LOCUS I26234 14 bp DNA PAT 07-OCT-1996
DEFINITION Sequence 19 from patent US 5556955.
ACCESSION I26234
VERSION I26234.1 GI:1606104
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 14)
AUTHORS Verghnaud, G.
TITLE Process for detection of new polymorphic loci in a DNA sequence, nucleotide sequences forming hybridization probes and their applications
JOURNAL Patent: US 5556955-A 19 17-SEP-1996;
FEATURES
source 1..14 Location/Qualifiers
BASE COUNT 4 a 4 c 2 g 4 t
ORIGIN

Query Match 100.0%; Score 7; DB 81; Length 14;
Best Local Similarity 100.0%; Pred. No. 5e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Search completed: March 27, 2001, 08:17:04
Job time: 5901 sec

Qy 1 tgttggg 7
Db 12 TGTGGG 6

RESULT 14
AX007732/c
LOCUS AX007732 15 bp DNA VRL 06-SEP-2000
DEFINITION Sequence 274 from Patent WO967428.
ACCESSION AX007732
VERSION AX007732.1 GI:9995429
KEYWORDS
SOURCE Aids-associated retrovirus.
ORGANISM Aids-associated retrovirus
REFERENCE 1 (bases 1 to 15)
AUTHORS Stuyver, L.
TITLE Method for detection of drug-selected mutations in the hiv protease gene
JOURNAL Patent: WO 967428-A 29-DEC-1999;
INNOGENETICS NV (BE); STUYVER LIEVEN (BE)
FEATURES
source
1..15
/organism="Aids-associated retrovirus"
/db_xref="taxon:11966"
BASE COUNT 6 a 6 c 0 g 3 t
ORIGIN

Query Match 100.0%; Score 7; DB 79; Length 15;
Best Local Similarity 100.0%; Pred. No. 5e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgttggg 7
Db 13 TGTGGG 7

RESULT 15
A88272/c
LOCUS A88272 15 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 420 from Patent WO9833904.
ACCESSION A88272
VERSION A88272.1 GI:6736842
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified
REFERENCE 1 (bases 1 to 15)
AUTHORS Brysch, W. and Schlingensiepen, K.
TITLE AN ANTISENSE OLIGONUCLEOTIDE PREPARATION METHOD
JOURNAL Patent: WO 9833904-A 06-AUG-1998;
BIOGNOSTIK GES (DE); BRYSCH WOLFGANG (DE)
FEATURES
source
1..15
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 6 a 4 c 0 g 5 t
ORIGIN

Query Match 100.0%; Score 7; DB 81; Length 15;
Best Local Similarity 100.0%; Pred. No. 5e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgttggg 7
Db 12 TGTGGG 6

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 27, 2001, 08:22:58 ; Search time 207.51 Seconds
(without alignments)
12.672 Million cell updates/sec

Title: US-09-380-826A-4

Perfect score: 7

Sequence: 1 tgttga 7

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 18781343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_36:*

```
1: /cgn2_2/gcgdata/geneseq/geneseq/NA1980.DAT:*
2: /cgn2_2/gcgdata/geneseq/geneseq/NA1981.DAT:*
3: /cgn2_2/gcgdata/geneseq/geneseq/NA1982.DAT:*
4: /cgn2_2/gcgdata/geneseq/geneseq/NA1983.DAT:*
5: /cgn2_2/gcgdata/geneseq/geneseq/NA1984.DAT:*
6: /cgn2_2/gcgdata/geneseq/geneseq/NA1985.DAT:*
7: /cgn2_2/gcgdata/geneseq/geneseq/NA1986.DAT:*
8: /cgn2_2/gcgdata/geneseq/geneseq/NA1987.DAT:*
9: /cgn2_2/gcgdata/geneseq/geneseq/NA1988.DAT:*
10: /cgn2_2/gcgdata/geneseq/geneseq/NA1989.DAT:*
11: /cgn2_2/gcgdata/geneseq/geneseq/NA1990.DAT:*
12: /cgn2_2/gcgdata/geneseq/geneseq/NA1991.DAT:*
13: /cgn2_2/gcgdata/geneseq/geneseq/NA1992.DAT:*
14: /cgn2_2/gcgdata/geneseq/geneseq/NA1993.DAT:*
15: /cgn2_2/gcgdata/geneseq/geneseq/NA1994.DAT:*
16: /cgn2_2/gcgdata/geneseq/geneseq/NA1995.DAT:*
17: /cgn2_2/gcgdata/geneseq/geneseq/NA1996.DAT:*
18: /cgn2_2/gcgdata/geneseq/geneseq/NA1997.DAT:*
19: /cgn2_2/gcgdata/geneseq/geneseq/NA1998.DAT:*
20: /cgn2_2/gcgdata/geneseq/geneseq/NA1999.DAT:*
21: /cgn2_2/gcgdata/geneseq/geneseq/NA2000.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	7	100.0	7	19	V58898
2	7	100.0	10	20	X99938
3	7	100.0	10	21	Z78251
4	7	100.0	10	21	Z80950
5	7	100.0	10	21	Z82660
6	7	100.0	10	21	Z84297
7	7	100.0	11	20	X77658
8	7	100.0	14	15	Q40606
9	7	100.0	14	15	Q78406
10	7	100.0	14	19	V97198
11	7	100.0	14	19	V48830
12	7	100.0	14	19	V48832

13	7	100.0	14	20	A21661	Integrin alpha 6 s
14	7	100.0	14	20	V92044	Human C-raf target
15	7	100.0	14	21	Z97785	HIV-1 protease gen
16	7	100.0	14	21	Z97786	HIV-1 protease gen
17	7	100.0	14	21	Z65483	Immunosuppressant
18	7	100.0	14	21	Z64788	Substrate for hair
19	7	100.0	15	2	N10044	Sequence of initia
20	7	100.0	15	2	N10045	Sequence of initia
21	7	100.0	15	17	X66645	Human CD40 hammerh
22	7	100.0	15	17	X66275	Mouse B7-2 hammerh
23	7	100.0	15	17	X66276	Mouse B7-2 hammerh
24	7	100.0	15	17	X66274	Mouse B7-2 hammerh
25	7	100.0	15	17	X66235	Mouse B7-2 hammerh
26	7	100.0	15	17	X66236	Mouse B7-2 hammerh
27	7	100.0	15	17	X66237	Mouse B7-2 hammerh
28	7	100.0	15	17	X66238	Mouse B7-2 hammerh
29	7	100.0	15	17	X65872	Human B7-2 hammerh
30	7	100.0	15	17	X65873	Human B7-2 hammerh
31	7	100.0	15	17	Z50248	Rabbit CERP HH rib
32	7	100.0	15	17	Z50250	Rabbit CERP HH rib
33	7	100.0	15	19	V49025	rb gene antisense
34	7	100.0	15	19	V48831	ErbB-2 gene antisense
35	7	100.0	15	20	X31673	Tag sequence of a
36	7	100.0	15	21	Z59285	Human NR8 gene pro
37	7	100.0	15	21	Z59298	Human NR8 gene pro
38	7	100.0	15	21	Z59299	Human NR8 gene pro
39	7	100.0	15	21	Z90823	Human NR8 gene pro
40	7	100.0	15	21	Z90839	Human NR8 gene pro
41	7	100.0	15	21	Z90896	Human NR8 gene pro
42	7	100.0	15	21	Z90905	Human NR8 gene pro
43	7	100.0	15	21	Z97784	HIV-1 protease gen
44	7	100.0	15	21	Z62590	Substrate for HH r
45	7	100.0	15	21	Z62635	Substrate for HH r

ALIGNMENTS

RESULT 1

```
V58898
ID V58898 standard; DNA; 7 BP.
XX
AC V58898;
XX
DT 20-JAN-1999 (first entry)
XX
DE Leptospira nucleotide sequence.
XX
KW Infection; pathogenic Leptospira; protective immunity; therapy;
diagnosis; ss.
XX
OS Leptospira sp.
XX
PN WO9840099-A1.
XX
PD 17-SEP-1998.
XX
PF 06-MAR-1998; 98WO-AU00145.
XX
PR 07-MAR-1997; 97AU-0005494.
XX
PA (AGRI-) AGRIC VICTORIA SERVICES PTY LTD.
PI (PIGR-) PIC RES & DEV CORP.
XX
Chappel RJ;
XX
WPI; 1998-520791/44.
XX
New isolated pathogenic Leptospira bacterium - useful for, e.g
developing products for conferring protective immunity, and for
prophylactic or therapeutic treatment
XX
Claim 15; Page 71; 94pp; English.
PS
```

XX This sequence represents a *Leptospira* DNA sequence isolated from the
 CC pathogenic *Leptospira* (LS) bacterium of the invention. The bacterium
 CC belongs to serogroup Hurstbridge or serovar Hurstbridge or the species
 CC *L. fainei*. The LS bacteria can be used for conferring protective
 CC immunity against pathogenic LS bacteria in humans or animals. The
 CC bacteria can also be used for prophylactic or therapeutic treatment of LS
 CC infections. The DNAs and antibodies may also be used for detection and
 CC diagnosis of past or present LS infection.
 XX
 SQ Sequence 7 BP; 1 A; 0 C; 3 G; 3 T; 0 other;

Query Match 100.0%; Score 7; DB 19; Length 7;
 Best Local Similarity 100.0%; Pred. No. 5.2e+07;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgttggg 7
 Db 1 tgttggg 7
 |||||

RESULT 2
 X9938/c
 ID X9938 standard; DNA; 10 BP.

XX AC X9938;
 XX 21-OCT-1999 (first entry)
 DT Human parkin gene intron 7 fragment.

DE Parkinson's disease related gene; parkin gene; variant; gene therapy;
 KW intron; ss.

XX OS Homo sapiens.

XX WO9940191-A1.

XX PD 12-AUG-1999.

XX 09-FEB-1999; 99WO-JP00545.

XX 09-FEB-1998; 98JP-0027531.

XX (MIZU)/ MIZUNO Y.
 XX (SHIM)/ SHIMIZU N.

XX Mizuno Y, Shimizu N;

XX WPI; 1999-494295/41.

XX Gene implicated in the pathology of Parkinson's disease, used for
 PT treatment of the disease

XX Claim 11; Page 100; 114pp; English.

XX This sequence represents a fragment of an intron from the gene of the
 CC invention. The gene has been designated the parkin gene, and variants of
 CC it are implicated in the pathology of Parkinson's disease, and found in
 CC Parkinson's disease patients. The sequences may be used for the
 CC diagnosis, treatment (including gene therapy) and investigation of
 CC Parkinson's disease.

XX Sequence 10 BP; 3 A; 3 C; 1 G; 3 T; 0 other;

Query Match 100.0%; Score 7; DB 20; Length 10;
 Best Local Similarity 100.0%; Pred. No. 1.4e+04;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgttggg 7
 |||||

Db 9 TGTGGA 3
 RESULT 3
 ID Z78251 standard; DNA; 10 BP.
 XX Z78251;
 AC
 XX
 DT 10-APR-2000 (first entry)
 DE Human dendritic cell SAGE tag, SEQ ID NO:679.

XX SAGE tag: serial analysis of gene expression; antigen-presenting cell;
 KW APC; monocyte-derived dendritic cell; differential gene expression;
 KW immunostimulatory cofactor; costimulatory factor; CTL;
 KW cytotoxic T-lymphocyte; tumour antigen; immunotherapy; anticancer; ss.

XX OS Homo sapiens.

XX WO9965924-A2.

XX PD 23-DEC-1999.

XX PF 18-JUN-1999; 99WO-US13800.

XX 19-JUN-1998; 98US-0089833.

XX 19-JUN-1998; 98US-0089844.

XX 19-JUN-1998; 98US-0089853.

XX 19-JUN-1998; 98US-0089878.

XX 19-JUN-1998; 98US-0089991.

XX 19-JUN-1998; 98US-0089992.

XX 19-JUN-1998; 98US-0089993.

XX 19-JUN-1998; 98US-0089994.

XX 19-JUN-1998; 98US-0089997.

XX 19-JUN-1998; 98US-0089999.

XX 19-JUN-1998; 98US-0090000.

XX 19-JUN-1998; 98US-0090035.

XX 19-JUN-1998; 98US-0090036.

XX 19-JUN-1998; 98US-0090039.

XX 19-JUN-1998; 98US-0090040.

XX 19-JUN-1998; 98US-0090041.

XX 19-JUN-1998; 98US-0090042.

XX 19-JUN-1998; 98US-0090043.

XX 19-JUN-1998; 98US-0090044.

XX 19-JUN-1998; 98US-0090045.

XX 19-JUN-1998; 98US-0090047.

XX 19-JUN-1998; 98US-0090072.

XX 19-JUN-1998; 98US-0090076.

XX 19-JUN-1998; 98US-0090077.

XX 19-JUN-1998; 98US-0090078.

XX 19-JUN-1998; 98US-0090079.

XX 19-JUN-1998; 98US-0090080.

XX 08-DEC-1998; 98US-0111715.

(GENZ) GENZYME CORP.

PA (ROBE/) ROBERTS B L.

PA (SHAN/) SHANKARA S.

PI Roberts BL, Shankara S;

XX WPI; 2000-106077/09.

XX Isolated polynucleotides differentially expressed in antigen-presenting
 PT cells, useful in gene vaccines against cancer -

XX Claim 1; Page 84; 130pp; English.

XX Sequences Z77573-Z79709 represent SAGE (serial analysis of gene
 CC expression) tags used to identify mRNA transcripts encoding
 CC immunostimulatory cofactor proteins which are preferentially or
 CC differentially expressed in monocyte-derived dendritic cells compared

CC with monocytes. Some of the transcripts correspond to known genes or
 CC ESTs (expressed sequence tags) which were previously unknown to be
 CC preferentially or differentially expressed in dendritic cells, while
 CC other transcripts correspond to novel genes. Antigen-presenting cell
 CC (APC)-associated costimulatory factors play an important role in the
 CC activation of the cytotoxic immune response, particularly against tumour
 CC cells. Tumour antigen presentation via the MHC (major histocompatibility
 CC complex) and subsequent recognition by T-cell receptors is alone
 CC insufficient to activate a robust cytotoxic immune response that can lyse
 CC the tumour cells, immunostimulatory cofactors also being required for
 CC efficient activation of cytotoxic T-lymphocytes (CTLs). Nucleic acid
 CC sequences identified using the SAGE tags have several potential uses.
 CC They may be used in vaccines to induce an immune response, particularly
 CC against a tumour antigen; to modulate the genotype of an APC; to screen
 CC for agents that modulate expression of differentially expressed genes in
 CC an APC; and as hybridisation probes/amplification primers for the
 CC diagnosis, prognosis and monitoring of diseases related to abnormal
 CC expression of these genes. Detection of the dendritic cell differentially
 CC expressed genes, or of their encoded proteins, can be used to identify
 CC cells as belonging to the monocyte lineage. Cells containing these genes
 CC can be used in active immunotherapy (or to stimulate production of a
 CC population of antigen-specific effector cells) and vectors containing
 CC them are used in gene therapy. Co-administration of tumour antigens and
 CC APC-associated costimulatory factors ensures adequate antigen
 CC presentation to endogenous APCs and upregulates the APCs for the
 CC presentation of co-stimulatory signals, migration to T cell-rich sites,
 CC secretion of T cell growth factors and secretion of chemokines for
 CC recruitment of immune effector cells.
 CC
 CC Sequence 10 BP; 1 A; 2 C; 3 G; 4 T; 0 other;

Query Match 100.0%; Score 7; DB 21; Length 10;
 Best Local Similarity 100.0%; Pred. No. 1.4e+04;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgttggg 7
 |||||
 Db 3 tgttggg 9

RESULT 4
 Z80950
 ID Z80950 standard; DNA; 10 BP.
 AC Z80950;

07-APR-2000 (first entry)

DE Metastatic breast tumour cell upregulated transcript tag #184.

KW Human; metastatic breast tumour tissue; breast cancer; tag; primer;
 KW non-metastatic breast tumour tissue; gene therapy; anticancer;
 KW antimetastatic; vaccine; diagnosis; ss.

OS Homo sapiens.

XX WO9965928-A2.

PN 23-DEC-1999.

XX 18-JUN-1999; 99WO-US13647.

PF 19-JUN-1998; 98US-0089853.

PR 19-JUN-1998; 98US-0089997.

PR 19-JUN-1998; 98US-0090039.

PR 19-JUN-1998; 98US-0090040.

PR 19-JUN-1998; 98US-0090041.

XX (GENZ) GENZYME CORP.

PA (ROBE/) ROBERTS B L.

PA (SHAN/) SHANKARA S.

PI Roberts BL, Shankara S;

XX WPI; 2000-106079/09.

XX Isolated polynucleotides differentially expressed between metastatic
 PT and non-metastatic breast cancer cells, useful for diagnosis,
 PT prevention and treatment of cancer -

XX Claim 1; Page 63; 219pp; English.

XX Z80767 to Z83941 represent tags corresponding to distinct transcripts
 CC that are preferentially transcribed in the metastatic breast tumour
 CC tissue (i.e. are upregulated in Metastatic breast tumour cells). Z83942
 CC to Z86677 represent tags corresponding to distinct transcripts that are
 CC preferentially transcribed in the primary or non-metastatic breast tumour
 CC tissue (i.e. are downregulated in metastatic breast tumour cells).
 CC These transcripts can be used for diagnosis, prognosis, monitoring and
 CC treatment of breast cancer, particularly where metastatic. Diagnosis is
 CC by standard immunoassays or hybridisation/amplification reactions.
 CC Compounds that modulate expression of the transcripts are potentially
 CC useful for treatment of (metastatic) breast cancer, while promoters from
 CC the transcripts are used to direct expression, in selected cell types,
 CC of e.g. therapeutic genes (also ribozymes or antisense sequences),
 CC particularly an antigen-encoding sequence for use in gene or cell-based
 CC vaccines. Polypeptides encoded by the transcripts are also useful in
 CC vaccines; for diagnosing breast cancer and for raising specific
 CC antibodies (Ab). Ab are used to detect the polypeptides or as therapeutic
 CC agents. Host cells that produce the polypeptides can be used to expand
 CC and isolate populations of educated, antigen-specific immune effector
 CC cells, e.g. cytotoxic T lymphocytes, and these used for adoptive
 CC immunotherapy.

XX Sequence 10 BP; 2 A; 0 C; 4 G; 4 T; 0 other;

Query Match 100.0%; Score 7; DB 21; Length 10;
 Best Local Similarity 100.0%; Pred. No. 1.4e+04;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgttggg 7
 |||||
 Db 4 tgttggg 10

RESULT 5
 Z82660
 ID Z82660 standard; DNA; 10 BP.
 AC Z82660;

07-APR-2000 (first entry)

DE Metastatic breast tumour cell upregulated transcript tag #1894.

KW Human; metastatic breast tumour tissue; breast cancer; tag; primer;
 KW non-metastatic breast tumour tissue; gene therapy; anticancer;
 KW antimetastatic; vaccine; diagnosis; ss.

OS Homo sapiens.

XX WO9965928-A2.

PN 23-DEC-1999.

XX 18-JUN-1999; 99WO-US13647.

PF 19-JUN-1998; 98US-0089853.

PR 19-JUN-1998; 98US-0089997.

PR 19-JUN-1998; 98US-0090039.

PR 19-JUN-1998; 98US-0090040.

PR 19-JUN-1998; 98US-0090041.

XX (GENZ) GENZYME CORP.

```

PA (ROBE/) ROBERTS B L.
PA (SHAN/) SHANKARA S.
PI Roberts BL, Shankara S;
XX WPI: 2000-106079/09.
XX
XX Isolated polynucleotides differentially expressed between metastatic
XX and non-metastatic breast cancer cells, useful for diagnosis,
XX prevention and treatment of cancer -
XX Claim 1: Page 110; 219pp; English.
XX
XX 280767 to 283941 represent tags corresponding to distinct transcripts
XX that are preferentially transcribed in the metastatic breast tumour
XX tissue (i.e. are upregulated in metastatic breast tumour cells). 283942
XX to 286677 represent tags corresponding to distinct transcripts that are
XX preferentially transcribed in the primary or non-metastatic breast tumour
XX tissue (i.e. are downregulated in metastatic breast tumour cells).
XX These transcripts can be used for diagnosis, prognosis, monitoring and
XX treatment of breast cancer, particularly where metastatic. Diagnosis is
XX by standard immunoassays or hybridisation/amplification reactions.
XX Compounds that modulate expression of the transcripts are potentially
XX useful for treatment of (metastatic) breast cancer, while promoters from
XX the transcripts are used to direct expression, in selected cell types,
XX of e.g. therapeutic genes (also ribozymes or antisense sequences),
XX particularly an antigen-encoding sequence for use in gene or cell-based
XX vaccines. Polypeptides encoded by the transcripts are also useful in
XX vaccines; for diagnosing breast cancer and for raising specific
XX antibodies (Ab). Ab are used to detect the polypeptides or as therapeutic
XX agents. Host cells that produce the polypeptides can be used to expand
XX and isolate populations of educated, antigen-specific immune effector
XX cells, e.g. cytotoxic T lymphocytes, and these used for adoptive
XX immunotherapy.
XX Sequence 10 BP; 2 A; 0 C; 4 G; 4 T; 0 other;

PR 19-JUN-1998; 98US-0090041.
XX (GENZ ) GENZYME CORP.
PA (ROBE/) ROBERTS B L.
PA (SHAN/) SHANKARA S.
XX
XX Roberts BL, Shankara S;
XX WPI: 2000-106079/09.
XX
XX Isolated polynucleotides differentially expressed between metastatic
XX and non-metastatic breast cancer cells, useful for diagnosis,
XX prevention and treatment of cancer -
XX Claim 1: Page 153; 219pp; English.
XX
XX 280767 to 283941 represent tags corresponding to distinct transcripts
XX that are preferentially transcribed in the metastatic breast tumour
XX tissue (i.e. are upregulated in metastatic breast tumour cells). 283942
XX to 286677 represent tags corresponding to distinct transcripts that are
XX preferentially transcribed in the primary or non-metastatic breast tumour
XX tissue (i.e. are downregulated in metastatic breast tumour cells).
XX These transcripts can be used for diagnosis, prognosis, monitoring and
XX treatment of breast cancer, particularly where metastatic. Diagnosis is
XX by standard immunoassays or hybridisation/amplification reactions.
XX Compounds that modulate expression of the transcripts are potentially
XX useful for treatment of (metastatic) breast cancer, while promoters from
XX the transcripts are used to direct expression, in selected cell types,
XX of e.g. therapeutic genes (also ribozymes or antisense sequences),
XX particularly an antigen-encoding sequence for use in gene or cell-based
XX vaccines. Polypeptides encoded by the transcripts are also useful in
XX vaccines; for diagnosing breast cancer and for raising specific
XX antibodies (Ab). Ab are used to detect the polypeptides or as therapeutic
XX agents. Host cells that produce the polypeptides can be used to expand
XX and isolate populations of educated, antigen-specific immune effector
XX cells, e.g. cytotoxic T lymphocytes, and these used for adoptive
XX immunotherapy.
XX Sequence 10 BP; 2 A; 0 C; 3 G; 5 T; 0 other;

Query Match 100.0%; Score 7; DB 21; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.4e+04;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 tggttggg 7
Db 2 tggttggg 8

RESULT 7
X77658/c
ID X77658 standard; DNA; 11 BP.
XX
XX X77658;
XX
XX 09-AUG-1999 (first entry)
XX
XX N11 active EGS 22.
XX
XX External guide sequence; EGS; target mRNA; identification; diagnostic;
XX inactivation; essential gene; therapy; ss.
XX Synthetic.
XX
XX WO9927135-A2.
XX
XX 03-JUN-1999.
XX
XX 20-NOV-1998; 98WO-US24854.
XX
XX 30-MAR-1998; 98US-0079851.
XX
XX 21-NOV-1997; 97US-0976220.
XX

```


XX (INNO-) INNOVIR LAB INC.
 XX Kindt TJ, Nilsen TW, Robertson HD;
 XX WPI; 1999-357853/30.
 DR
 XX Identifying and inhibiting functional nucleic acid molecules in
 PT cells
 PT
 XX Example 3; Page 28; 58pp; English.
 PS
 XX This invention describes a novel method allowing essential or functional
 CC genes to be rapidly identified and inactivated. The method is able to
 CC firstly identify most of the essential genes in an organism (i.e. a
 CC bacteria or a eukaryote) needed for survival, and secondly it provides
 CC for reducing or inactivating their expression. The method is able to
 CC identify functional oligonucleotide molecules able to be used as
 CC diagnostic reagents and therapeutics. The method provides a means for
 CC identifying essential genes whose sequence is known only as part of a
 CC genome with unknown function, as well as a means for identifying
 CC functional oligonucleotide molecules. The method involves the use of a
 CC nucleic acid molecule comprising (a) a first reporter gene encoding a
 CC fusion protein comprising a protein of interest (itself translated from
 CC an RNA of interest) and a reporter protein, a second reporter gene
 CC encoding a second reporter protein, and (c) a targeting gene encoding a
 CC functional oligonucleotide molecule such as an external guide sequence
 CC (EGS), a ribozyme or an antisense RNA and targeted to the RNA of interest
 CC at a site on the first reporter gene able to encode the RNA of interest.
 XX
 XX Sequence 11 BP; 4 A; 4 C; 1 G; 2 T; 0 other;
 SQ

Query Match 100.0%; Score 7; DB 20; Length 11;
 Best Local Similarity 100.0%; Pred. No. 1.4e+04;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 tgtttgga 7
 Db 10 TGTGTGGA 4

RESULT 8
 Q40606/C
 ID Q40606 standard; DNA; 14 BP.
 XX
 AC Q40606;
 XX
 DT 10-AUG-1993 (first entry)
 XX
 DE Hypervariable region detection probe 14C12.
 XX
 KW HVR; human; animal; forensic science; paternity testing; diagnosis;
 KW animal breeding; hereditary diseases; tumours; allele; loss;
 KW chromosomal regions; tumour region identification; ss.
 XX
 OS Synthetic.
 XX
 PN FR2680520-A.
 XX
 PD 26-FEB-1993.
 XX
 PF 22-AUG-1991; 91FR-0010516.
 XX
 PR 22-AUG-1991; 91FR-0010516.
 XX
 PA (ETFR) ETAT FRANCAIS.
 XX
 PI Vergnaud G;
 XX
 DR WPI; 1993-136548/17.
 XX
 PT Detecting the hypervariable regions of DNA for diagnosing

PT hereditary illnesses and tumours - by hybridising labelled
 PT polynucleotides and analysing genomic DNA of individuals which
 PT react with restriction fragments
 XX
 PS Example; Page 13; 46pp; French.
 XX
 CC The sequence is that of a polynucleotide probe which may be used in
 CC the detection of new hypervariable regions (HVR) in a DNA sequence.
 CC HVR represent a fingerprint useful in e.g. forensic science,
 CC paternity testing, animal breeding, etc. The probe may be used as
 CC part of a method for the efficient detection in humans or other
 CC animals, without the use of mini-satellites or primary enrichment.
 XX
 XX Sequence 14 BP; 4 A; 4 C; 2 G; 4 T; 0 other;
 SQ

Query Match 100.0%; Score 7; DB 14; Length 14;
 Best Local Similarity 100.0%; Pred. No. 1.4e+04;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 tgtttgga 7
 Db 12 TGTGTGGA 6

RESULT 9
 Q78406
 ID Q78406 standard; DNA; 14 BP.
 XX
 AC Q78406;
 XX
 DT 27-JUN-1995 (first entry)
 XX
 DE Antisense oligonucleotide hybridising to TGF-beta gene.
 XX
 KW Transforming growth factor beta; TGF-beta; antisense; treatment;
 KW tumour; angiogenesis; breast tumour; neurofibroma; glioma;
 KW glioblastoma; carcinogenesis; carcinoma; oesophagus; oesophageal;
 KW gastric; gut; immunosuppression; oligonucleotide; ss.
 XX
 OS Synthetic.
 XX
 PN WO9425588-A.
 XX
 PD 10-NOV-1994.
 XX
 PF 29-APR-1994; 94WO-EP01362.
 XX
 PR 30-APR-1993; 93EP-0107089.
 PR 13-MAY-1993; 93EP-0107849.
 XX
 PA (BIOG-) BIOGNOSTIK GES BIOMOLEKULARE DIAGNOSTIK.
 XX
 PI Bogdahn U, Brysch W, Schlingensiepen G, Schlingensiepen K;
 PI Schlingensiepen R;
 XX
 DR WPI; 1994-358266/44.
 XX
 XX New transforming growth factor beta anti-sense
 PT oligo-nucleotide(s) - for treating immunosuppression, tumours,
 PT etc.
 XX
 PS Claim 6; Page 40; 74pp; English.
 XX
 CC The antisense oligonucleotides are useful in the treatment of
 CC tumours in which expression of TGF-beta is of relevance for
 CC pathogenicity and/or inhibition of pathological angiogenesis. They
 CC are used especially for the treatment of the immunosuppressive
 CC effect of TGF-beta, augmentation of the proliferation of cytotoxic
 CC lymphocytes, treatment of endogenous hyperexpression of TGF-beta,
 CC treatment of breast tumours, neurofibromas and malignant gliomas,
 CC including glioblastomas, treatment and prophylaxis of skin
 CC carcinogenesis, and treatment of oesophageal and gastric carcinomas.

Wed Mar 28 14:03:18 2001

CC ribozymes can be targeted to specific genes or to consensus sequences
 CC within a family of related genes, and being catalytic need to be
 CC present at only very low concentrations.
 XX
 SQ Sequence 14 BP; 4 A; 4 C; 5 G; 1 U; 0 other;

Query Match 100.0%; Score 7; DB 19; Length 14;
 Best Local Similarity 100.0%; Pred. No. 1.4e+04;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgttggg 7
 Db 13 TGTGGA 7

RESULT 11
 V48830/c
 ID V48830 standard; DNA; 14 BP.

XX V48830;

XX 15-OCT-1998 (first entry)

XX Erbb-2 gene antisense oligonucleotide Erbb-2-N-39.

XX Erbb-2; antisense oligonucleotide; modulate; gene expression; ss.

XX Synthetic.

XX Homo sapiens.

XX EP856579-A1.

XX 05-AUG-1998.

XX 31-JAN-1997; 97EP-0101531.

XX 31-JAN-1997; 97EP-0101531.

XX (BIOG-) BIOGNOSTIK GES BIOMOLEKULARE DIAGNOSTIK.

XX Brysch W, Schlingensiepen K;

XX WPI; 1998-400910/35.

XX Preparation of antisense oligo:nucleotide(s) which lack long runs of
 PT consecutive guanosine or inosine - and have specific ratio of
 PT residues able to form two or three hydrogen bonds, have greater
 PT activity and reduced toxicity, used therapeutically or to modulate
 PT growth of cells in culture

XX Example 4; Fig 6c; 286pp; English.

XX V48709-886 represent antisense oligonucleotides directed against the
 CC Erbb-2 gene. Of these, only oligonucleotides V48709-91 resulted
 CC in significant reduction in Erbb-2 protein expression, while
 CC oligonucleotides V48792-886 had little effect. The oligonucleotides
 CC exemplify the invention. The specification describes oligonucleotides
 CC that contain 8-30 nucleotides, which contain at most 8 nucleotides that
 CC can each form three hydrogen bonds to cytosine; do not contain four
 CC consecutive nucleotides able to form three H-bonds each to four
 CC consecutive cytosines; do not contain two sequences of three consecutive
 CC nucleotides each able to form three H-bonds to three consecutive
 CC cytosines, and the ratio between residues able to form two H-bonds each
 CC (2R) or three such bonds (3R) is given by 2R/3R = 0.33-0.72. The
 CC oligonucleotides are used to modulate expression of genes, particularly
 CC the genes for p53, ErB-2, junB, junD, TGF-beta 1 or beta 2 to control
 CC proliferation of primary cell cultures (e.g. bone marrow stem, liver or
 CC kidney cells, osteoclasts, osteoblasts and/or keratinocytes). The
 CC oligonucleotides can also be used to analyse function of proteins (by
 CC altering their expression or activity) and therapeutically, e.g. in cases
 CC of cancer or (targeting TGF) for stimulating the immune system.

XX

CC See 078352-Q78488. The sequences given in GENESQ files
 CC 078352-Q78407 and 078488 are antisense oligodeoxynucleotides of
 CC TGF-beta 1. The sequences given in GENESQ files 078408-78487 are
 CC antisense oligodeoxynucleotides of TGF-beta 2 in the form of
 CC phosphorothioate analogues.
 XX

SQ Sequence 14 BP; 3 A; 3 C; 4 G; 4 T; 0 other;

Query Match 100.0%; Score 7; DB 15; Length 14;
 Best Local Similarity 100.0%; Pred. No. 1.4e+04;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttttggg 7
 Db 3 ttttggg 9

RESULT 10
 V97198/c
 ID V97198 standard; RNA; 14 BP.

XX V97198;

XX 01-MAR-1999 (first entry)

XX Potato citrate synthase target sequence position 123.

XX Solanidine; glucosyltransferase; potato; citrate synthase; target;

XX hammerhead ribozyme; hairpin ribozyme; alkaloid biosynthesis;

XX flower formation; cleavage; solanaceous plant; ss.

XX Solanum tuberosum.

XX WO9832843-A2.

XX 30-JUL-1998.

XX 14-JAN-1998; 98WO-US00738.

XX 24-NOV-1997; 97US-0979416.

XX 28-JAN-1997; 97US-0036545.

XX 28-JAN-1997; 97US-0036599.

XX (RIBO-) RIBOZYME PHARM INC.

XX McSwiggen JA, Zwick MG;

XX WPI; 1998-427939/36.

XX New enzymatic nucleic acid(s) - useful for, e.g. reducing alkaloid

XX biosynthesis or regulating flowering

XX Claim 54; Page 59; 79pp; English.

XX The present invention describes enzymatic nucleic acid molecules with
 CC RNA-cleaving activity (e.g. ribozymes) which are capable of modulating
 CC the expression of plant genes: (i) involved in biosynthesis of
 CC alkaloids; or (ii) involved in flower formation. V95982 to V96334, and
 CC V96335 to V96354 represent potato solanidine glucosyltransferase
 CC hammerhead and hairpin ribozymes, respectively. V95629 to V95981, and
 CC V96355 to V96734 represent potato solanidine glucosyltransferase target
 CC sequences. V96773 to V97170, and V97171 to V97195 represent potato
 CC citrate synthase hammerhead and hairpin ribozymes, respectively. V96735
 CC to V96772, and V97196 to V97220 represent potato citrate synthase target
 CC sequences. Ribozymes of the present invention can be used to inhibit
 CC the synthesis of toxic alkaloids in solanaceous plants, particularly
 CC potato but also tomato, pepper, aubergine and ditura or to inhibit
 CC flowering in potato, lettuce, spinach, cabbage, brussel sprouts,
 CC arugula, kale, collards, chard, beet, turnip, sweet potato and turf
 CC grass. Also the ribozymes can be used for RNA manipulation in the same
 CC way that restriction endonucleases are for DNA, as well as to examine
 CC genetic drift and mutations in plants and to detect specific RNA. The

SQ Sequence 14 BP; 5 A; 4 C; 0 G; 5 T; 0 other;

Query Match 100.0%; Score 7; DB 19; Length 14;
Best Local Similarity 100.0%; Pred. No. 1.4e+04;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgttggg 7
DB 11 TGTGGGA 5

RESULT 12

V48832/c
ID V48832 standard; DNA; 14 BP.

XX AC
XX AC
XX AC

DT 15-OCT-1998 (first entry)

XX ErbB-2 gene antisense oligonucleotide ErbB-2-N-41.

DE ErbB-2; antisense oligonucleotide; modulate; gene expression; ss.

XX Synthetic.

OS Homo sapiens.

XX EP856579-A1.

PN 05-AUG-1998.

PD 31-JAN-1997; 97EP-0101531.

XX 31-JAN-1997; 97EP-0101531.

PR (BIOC-) BIOGNOSTIK GES BIOMOLEKULARE DIAGNOSTIK.

PA Brysch W, Schlingensiepen K;

PI WPI; 1998-400910/35.

XX Preparation of antisense oligo:nucleotide(s) which lack long runs of

PT consecutive cytosines or inosine - and have specific ratio of

PT residues able to form two or three hydrogen bonds, have greater

PT activity and reduced toxicity, used therapeutically or to modulate

PT growth of cells in culture

XX Example 4; Fig 6c; 286pp; English.

PS V48709-886 represent antisense oligonucleotides directed against the

XX ErbB-2 gene. Of these, only oligonucleotides V48709-91 resulted

CC in significant reduction in ErbB-2 protein expression, while

CC oligonucleotides V48792-886 had little effect. The oligonucleotides

CC exemplify the invention. The specification describes oligonucleotides

CC that contain 8-30 nucleotides, which contain at most 8 nucleotides that

CC can each form three hydrogen bonds to cytosine; do not contain four

CC consecutive nucleotides able to form three H-bonds each to four

CC consecutive cytosines; do not contain two sequences of three consecutive

CC nucleotides each able to form three H-bonds to three consecutive

CC cytosines, and the ratio between residues able to form two H-bonds each

CC (2R) or three such bonds (3R) is given by $2R/3R = 0.33-0.72$. The

CC oligonucleotides are used to modulate expression of genes, particularly

CC the genes for p53, ErbB-2, JunB, JunD, TGF-beta 1 or beta 2 to control

CC proliferation of primary cell cultures (e.g. bone marrow stem, liver or

CC kidney cells, osteoclasts, osteoblasts and/or keratinocytes). The

CC oligonucleotides can also be used to analyse function of proteins (by

CC altering their expression or activity) and therapeutically, e.g. in cases

CC of cancer or (targeting TGF) for stimulating the immune system.

XX Sequence 14 BP; 6 A; 4 C; 0 G; 4 T; 0 other;

SQ Query Match 100.0%; Score 7; DB 19; Length 14;

Best Local Similarity 100.0%; Pred. No. 1.4e+04;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgttggg 7

DB 12 TGTGGGA 6

RESULT 13

A21661

ID A21661 standard; RNA; 14 BP.

XX AC A21661;

XX 19-JUN-2000 (first entry)

DT Integrin alpha 6 subunit target site SEQ ID NO:4887.

DE Human; aryl hydrocarbon nuclear transport; ARNT; TIE-2; angiogenesis;

XX Integrin alpha 6 subunit; integrin subunit beta 3; hairpin ribozyme;

KW hammerhead ribozyme; angiogenic factor; cytostatic; antidiabetic;

KW ophthalmologic; antiinflammatory; antiarthritic; antipsoriatic; ARMD;

KW dermatological; RNA cleavage; cancer; diabetic retinopathy; arthritis;

KW age related macular degeneration; inflammation; neovascular glaucoma;

KW myopic degeneration; psoriasis; verruca vulgaris; angiofibroma;

KW tuberosus sclerosis; pot-wine stain; Sturge Weber syndrome;

KW Kippel-Trenaunay-Weber syndrome; Osler-Weber-Rendu syndrome; ss.

XX Homo sapiens.

XX WO9950403-A2.

PN 07-OCT-1999.

PD 24-MAR-1999; 99WO-US06507.

XX 27-MAR-1998; 98US-0079678.

PR (RIBO-) RIBOZYME PHARM INC.

XX Pavco PA, Roberts E, Jarvis T, Coeshott C, McSwiggen JA;

PI WPI; 1999-591315/50.

XX Novel ribozymes for modulating the synthesis, expression and/or

PT stability of an mRNA encoding an angiogenic factors

PT Claim 55; Page 219; 305pp; English.

PS The present invention describes enzymatic cleavage RNA molecules with

XX RNA cleaving activity, which specifically cleave RNA encoded by an aryl

CC hydrocarbon nuclear transporter (ARNT) gene, an integrin subunit beta 3

CC gene, an integrin alpha 6 subunit gene, or a Tie-2 gene. Al6775 to

CC Al7167 and Al7561 to Al7622 represent ribozyme sequences for ARNT, and

CC Al7168 to Al7560 and Al7623 to Al7684 represent their corresponding

CC target sequences; Al7685 to Al8385 and Al9087 to Al9154 represent

CC ribozyme sequences for Tie-2, and Al8386 to Al9086 and Al9155 to Al9222

CC represent their corresponding target sequences; Al9223 to A20361 and

CC A21501 to A21595 represent ribozyme sequences for integrin alpha 6

CC subunit, and A20362 to A21500 and A21596 to A21688 represent their

CC corresponding target sequences; A21689 to A22475 and A23263 to A23342

CC represent ribozyme sequence for integrin subunit beta 3, and A22476 to

CC A23262, A23343 to A23422 represent their corresponding target sequences.

CC The ribozymes of the invention are used for modulating the synthesis,

CC expression and/or stability of an mRNA encoding angiogenic factors,

CC especially ARNT, integrin subunit beta-3, integrin subunit alpha-6, or

CC Tie-2. They are especially used to treat cancer, diabetic retinopathy,

CC age related macular degeneration (ARMD), inflammation, and arthritis,

CC as well as neovascular glaucoma, myopic degeneration, psoriasis, verruca

CC vulgaris, angiofibroma of tuberous sclerosis, pot-wine stains, Sturge

CC Weber syndrome, Kippel-Trenaunay-Weber syndrome, Osler-Weber-Rendu

CC syndrome, and other syndromes and diseases related to the levels of

CC ARNT, Tie-2, integrin subunit alpha-6, or integrin subunit beta-3.

XX SQ Sequence 14 BP; 5 A; 1 C; 4 G; 4 U; 0 other;

Query Match 100.0%; Score 7; DB 20; Length 14;
Best Local Similarity 57.1%; Pred. No. 1.4e+04;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Oy 1 tgttggga 7
Db 5 uguuugga 11

RESULT 14
V92044/c
ID V92044 standard; RNA; 14 BP.
AC V92044;
XX
XX 18-FEB-1999 (first entry)
XX Human C-raf target sequence nucleotide position 1952.
DE Human; c-raf; A-raf; B-raf; hammerhead ribozyme; hairpin ribozyme;
KW target; substrate; catalyst; modulation; expression; Raf gene;
KW delivery; screening; identification; synthesis; deprotection;
KW purification; cancer; inflammation; psoriasis; non-hepatic ascites;
KW infection; genetic drift; restenosis; rheumatoid arthritis; ss.
XX
OS Homo sapiens.
XX
XX WO98050330-A2.
XX
PD 12-NOV-1998.
XX
XX 05-MAY-1998; 98WO-US09249.
XX
PR 19-DEC-1997; 97US-0068212.
PR 09-MAY-1997; 97US-0046059.
PR 09-JUN-1997; 97US-0049002.
PR 03-JUL-1997; 97US-0051718.
PR 22-AUG-1997; 97US-0056808.
PR 02-OCT-1997; 97US-0061321.
PR 02-OCT-1997; 97US-0061324.
PR 05-NOV-1997; 97US-0064866.
XX
XX (RIBO-) RIBOZYME PHARM INC.
XX
XX Beaudry A, Beigelman L, Bellon L, Burgin A, Jarvis T;
PI Karpeisky A, Kisich K, Matulic-Adamic J, McSwiggen JA;
PI Parry T, Reynolds M, Svedler D, Thompson J, Workman CT;
XX
XX WPI; 1999-009494/01.
DR
XX Identifying new catalytic nucleic acid that modulates selected
XX processes, especially ribozymes that cleave Raf RNA for treating
XX cancer, restenosis, and also new ribozymes and modified nucleoside
XX triphosphates used as antiviral agents and synthons
XX
XX Claim 179; Page 156; 259pp; English.
XX
XX A method has been developed for the identification of a nucleic acid
XX capable of modulating a process in a biological system. The method
XX comprises: (a) introducing into the system a random library of nucleic
XX acid catalysts (NAC) having a substrate binding domain (SBD), comprising
XX a random sequence, and a catalytic domain (CD); and (b) identifying NAC
XX in systems where modulation has occurred and/or determining the sequence
XX of at least part of the SBDs in such systems. Nucleic acid molecules with
XX endonuclease activity and catalytic activity, from the present invention,
XX are used to modulate gene expression in plant and mammalian cells and to
XX cleave target nucleic acid, particularly for treating systemic diseases
XX caused by specific RNA, e.g. cancer, inflammation, psoriasis, non-hepatic
XX ascites and infection. They may also be used to detect genetic drift and

CC mutations in diseased cells and to determine c-raf RNA. Specifically NACs
CC with RNA-cleaving activity that modulate expression of the Raf gene, are
CC used to treat cancer, restenosis, psoriasis or rheumatoid arthritis, or
CC generally any condition associated with the level of c-raf. Introduction
CC of sugar/phosphate modifications increases stability against nuclease
CC and activity. V90922 to V93877 represent NACs that can be used in the
CC method, specifically for modulating the expression of a Raf gene.
XX
SQ Sequence 14 BP; 4 A; 5 C; 3 G; 2 U; 0 other;

Query Match 100.0%; Score 7; DB 20; Length 14;
Best Local Similarity 100.0%; Pred. No. 1.4e-04;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 tgttggga 7
Db 14 TGTGGGA 8

RESULT 15
297785/c
ID 297785 standard; DNA; 14 BP.
XX
XX 297785;
XX
DT 26-APR-2000 (first entry)
XX
XX HIV-1 protease gene probe SEQ ID NO:275.
XX
XX Human immunodeficiency virus; HIV; protease; probe; detection;
KW drug selected mutation; hybridisation; genotyping; infection;
KW drug resistance; ss.
XX
XX Human immunodeficiency virus type 1.
XX
XX WO9967428-A2.
XX
PD 29-DEC-1999.
XX
XX 22-JUN-1999; 99WO-EF04317.
XX
XX 24-JUN-1998; 98EP-0870143.
XX
XX (INNO-) INNOGENETICS NV.
XX
XX Stuyver L;
XX
XX WPI; 2000-147219/13.
XX
XX Detection of drug-selected mutations in the HIV protease gene used to
PT treat HIV infections
XX
XX Claim 3; Page 40; 76pp; English.
XX
XX The present invention describes the detection of drug-selected mutations
XX in the HIV protease gene. The method of detection allows the simultaneous
XX characterisation of a range of codons involved in drug resistance using
XX sets of probes optimised to function together in a reverse-hybridisation
XX assay. 297517 to 297997 represent specifically claimed probes for use in
XX the assay, and 297479 to 297501 represent specifically claimed HIV
XX protease gene polymorphic nucleotide sequences. 297502 to 297515, and
XX 298004 to 298007, represent PCR primers for the HIV protease gene, and
XX 297516 represents an HIV protease probe used in an example from the
XX present invention. The method, probes and primers can be used for the
XX detection of drug-selected mutations in the HIV protease gene. The
XX method allows the simultaneous characterisation of a range of codons
XX involved in drug resistance. The method may also be used for HIV
XX protease genotyping assays. The probes are able to discriminate between
XX wild type and mutated protease sequences. The method allows rapid and
XX reliable detection of drug-selected mutation in HIV.
XX
XX Sequence 14 BP; 5 A; 6 C; 0 G; 3 T; 0 other;

Query Match 100.0%; Score 7; DB 21; Length 14;
Best Local Similarity 100.0%; Pred. No. 1.4e+04;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ttttggga 7
 |||||
Db 12 TcTtTGGG 6

Search completed: March 27, 2001, 08:22:59
Job time: 5301 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 27, 2001, 08:17:06 ; Search time 2286.42 Seconds
(without alignments)
49.243 Million cell updates/sec

Title: US-09-380-826A-7
Perfect score: 22
Sequence: 1 tgttgatcacagatttgata 22

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 118133 seqs, 2558875100 residues

Total number of hits satisfying chosen parameters: 2236266

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_ph:*
6: gb_pl1:*
7: gb_pl2:*
8: gb_pr1:*
9: gb_pr2:*
10: gb_pr3:*
11: gb_ro:*
12: gb_sy:*
13: gb_uni:*
14: em_fun:*
15: em_hum1:*
16: em_hum2:*
17: em_in:*
18: em_om:*
19: em_or:*
20: em_ov:*
21: em_pat:*
22: em_ph:*
23: em_pl:*
24: em_ro:*
25: em_sy:*
26: em_uni:*
27: em_v1:*
28: em_v2:*
29: gb_ba3:*
30: gb_in1:*
31: gb_in2:*
32: gb_in3:*
33: gb_pl3:*
34: gb_pr4:*
35: em_ba1:*
36: em_ba2:*
37: em_htg1:*
38: em_htg2:*
39: em_htg3:*
40: em_htg4:*
41: em_htg5:*
42: em_htg6:*
43: em_htg7:*

44: em_htg8:*
45: em_htg9:*
46: em_htg10:*
47: em_hum3:*
48: em_hum4:*
49: em_hum5:*
50: em_hum6:*
51: gb_pr5:*
52: gb_pr6:*
53: gb_pr7:*
54: gb_htg1:*
55: gb_htg2:*
56: gb_htg3:*
57: gb_htg4:*
58: gb_htg5:*
59: gb_htg6:*
60: gb_htg7:*
61: gb_htg8:*
62: gb_htg9:*
63: gb_htg10:*
64: gb_htg11:*
65: gb_htg12:*
66: gb_htg13:*
67: gb_htg14:*
68: gb_htg15:*
69: gb_htg16:*
70: gb_htg17:*
71: gb_htg18:*
72: gb_htg19:*
73: gb_htg20:*
74: gb_htg21:*
75: gb_htg22:*
76: gb_htg23:*
77: gb_sts1:*
78: gb_sts2:*
79: gb_v11:*
80: gb_v12:*
81: gb_pat1:*
82: gb_pat2:*
83: em_htg0:*
84: gb_htg24:*
85: gb_pr8:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	22	100.0	1481	2	LFU60594	U60594 Leptosira
2	20.4	92.7	354	1	AB007014	AB007014 Spirochae
3	18.8	85.5	283	1	AF003953	AF003953 Leptospir
4	18.8	85.5	288	2	LIU94974	U94974 Leptosira
5	18.8	85.5	288	2	LIU94975	U94975 Leptosira
6	18.8	85.5	288	2	LIU94976	U94976 Leptosira
7	18.8	85.5	288	2	LIU94977	U94977 Leptosira
8	18.8	85.5	288	2	LIU94978	U94978 Leptosira
9	18.8	85.5	288	2	LIU94979	U94979 Leptosira
10	18.8	85.5	353	1	AB007012	AB007012 Spirochae
11	18.8	85.5	1012	7	ATHRS2X	M98336 Arabidopsis
12	18.8	85.5	72590	6	AC003000	AC003000 Arabidops
13	18.8	85.5	109723	72	AF165146	AF165146 Homo sapi
14	18.8	85.5	200799	6	AF137379	AF137379 Nephrosel
15	18.8	85.5	200799	6	AF137379	AF137379 Nephrosel
16	18.4	83.6	1812	31	AF220067	AF220067 Drosophil
17	18.4	83.6	27365	58	AC015209	AC015209 Drosophil
18	18.4	83.6	185087	54	AC007475	AC007475 Drosophil
19	18.4	83.6	262731	30	AE003823	AE003823 Drosophil
20	17.8	80.9	1874	79	EBBPC44NCP	L48441 Echinochloa
21	17.8	80.9	31214	33	SPCC830	AL109850 S.pombe c

Wed Mar 28 14:03:26 2001

```

22 17.8 80.9 85992 6 AB009052 Arabidops
23 17.8 80.9 91894 65 AC026327
24 17.8 80.9 174766 65 AC026320 Homo sapi
25 17.8 80.9 270889 67 AC055742
26 17.4 79.1 59261 33 T12M4
27 17.4 79.1 81662 6 AB008265
28 17.4 79.1 109512 68 AC068144
29 17.4 79.1 149241 61 AC021015
30 17.4 79.1 187847 62 AC023155
31 17.2 78.2 462 1 AF175672 Unculture
32 17.2 78.2 835 7 AF271231
33 17.2 78.2 1083 1 AF018567
34 17.2 78.2 2213 33 LEPA000ALB
35 17.2 78.2 3608 33 SCYL0031C
36 17.2 78.2 5596 32 SP040832
37 17.2 78.2 34496 31 CELR11G1
38 17.2 78.2 38655 31 CELC02F12
39 17.2 78.2 49311 84 F26K10
40 17.2 78.2 62478 71 AC083790
41 17.2 78.2 67858 66 AC036233
42 17.2 78.2 68040 76 AP002082
43 17.2 78.2 97714 7 AFE25024
44 17.2 78.2 101679 7 AFI9B15
45 17.2 78.2 112467 52 HS0505B13

AB009052 Arabidops
AC026327 Homo sapi
AC026320 Homo sapi
AC055742 Homo sapi
AC03114 Arabidops
AB008265 Arabidops
AC068144 Homo sapi
AC021015 Homo sapi
AC023155 Homo sapi
AF175672 Unculture
AF271231 Albugo ca
AF018567 Unidentif
D49367 Lithospermu
Z73136 S.cerevisia
U40832 Strongyloce
U41016 Caenorhabd
U41345 Caenorhabd
AL049803 Arabidops
AC083790 Homo sapi
AC036233 Homo sapi
AP002082 Homo sapi
AL078469 Arabidops
Z59052 Human DNA s

10-SEP-1998
partial sequence.

ALIGNMENTS

RESULT 1
LFU60594 1481 bp DNA BCT 10-SEP-1998
LOCUS Leptospiira fainel 16S ribosomal RNA gene, partial sequence.
DEFINITION U60594
ACCESSION U60594.1 GI:1408219
VERSION
KEYWORDS
SOURCE
ORGANISM
Bacteria; Spirochaetales; Leptospiaceae; Leptospiira.
REFERENCE
1 (bases 1 to 1481)
AUTHORS Perolat,P., Chappel,R.J., Adler,B., Baranton,G., Bulach,D.M.,
Billinghurst,M.L., Letocart,M., Merien,F. and Serrano,M.S.
TITLE Leptospiira fainel sp. nov., isolated from pigs in Australia
JOURNAL Int. J. Syst. Bacteriol. 48 Pt 3, 851-858 (1998)
MEDLINE
REFERENCE
2 (bases 1 to 1481)
AUTHORS Adler,B., Chappel,R.J., Baranton,G., Bulach,D.M.,
Billinghurst,M.L., Letocart,M., Merien,F., Serrano,M.S. and
Perolat,P.
TITLE Direct Submission
JOURNAL Submitted (12-JUN-1996) Microbiology, Monash University, Wellington
RD., Clayton, VIC 3168, Australia
FEATURES
Location/Qualifiers
source
1..1481
/organism="Leptospiira fainel"
/strain="Hurstbridge"
/db_xref="taxon:48782"
<1..>1481
/product="16S ribosomal RNA"
BASE COUNT 391 a 335 c 439 g 314 t
ORIGIN
Query Match 100.0%; Score 22; DB 2; Length 1481;
Best Local Similarity 100.0%; Pred. No. 3.9; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 0;

QY 1 ttttgatcacaaagtattgata 22
|||||
Db 155 TGTGGATCACAAAGATTGATA 176

RESULT 2
LFU60594 1481 bp DNA BCT 10-SEP-1998
LOCUS Leptospiira fainel 16S ribosomal RNA gene, partial sequence.
DEFINITION U60594
ACCESSION U60594.1 GI:1408219
VERSION
KEYWORDS
SOURCE
ORGANISM
Bacteria; Spirochaetales; Leptospiaceae; Leptospiira.
REFERENCE
1 (bases 1 to 1481)
AUTHORS Perolat,P., Chappel,R.J., Adler,B., Baranton,G., Bulach,D.M.,
Billinghurst,M.L., Letocart,M., Merien,F. and Serrano,M.S.
TITLE Leptospiira fainel sp. nov., isolated from pigs in Australia
JOURNAL Int. J. Syst. Bacteriol. 48 Pt 3, 851-858 (1998)
MEDLINE
REFERENCE
2 (bases 1 to 1481)
AUTHORS Adler,B., Chappel,R.J., Baranton,G., Bulach,D.M.,
Billinghurst,M.L., Letocart,M., Merien,F., Serrano,M.S. and
Perolat,P.
TITLE Direct Submission
JOURNAL Submitted (12-JUN-1996) Microbiology, Monash University, Wellington
RD., Clayton, VIC 3168, Australia
FEATURES
Location/Qualifiers
source
1..1481
/organism="Leptospiira fainel"
/strain="Hurstbridge"
/db_xref="taxon:48782"
<1..>1481
/product="16S ribosomal RNA"
BASE COUNT 391 a 335 c 439 g 314 t
ORIGIN
Query Match 100.0%; Score 22; DB 2; Length 1481;
Best Local Similarity 100.0%; Pred. No. 3.9; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 0;

QY 1 ttttgatcacaaagtattgata 22
|||||
Db 155 TGTGGATCACAAAGATTGATA 176

```

```

AB007014 354 bp DNA BCT 13-OCT-1997
LOCUS Spirochaeta sp. 16S rRNA gene, partial sequence.
DEFINITION AB007014
ACCESSION AB007014.1 GI:2516255
VERSION 16S ribosomal RNA.
KEYWORDS Spirochaeta sp. (sub_species:Freshwater obligate oligotroph,
SOURCE strain:FO-95) DNA.
ORGANISM Spirochaeta sp.
Bacteria; Spirochaetales; Spirochaetaceae; Spirochaeta.
REFERENCE
1 (bases 1 to 354)
AUTHORS Shin,M.-S.
TITLE Direct Submission
JOURNAL Submitted (05-SEP-1997) to the DDBJ/EMBL/GenBank databases. Mi-Sun
Shin, Laboratory of Marine Molecular Microbiology, Faculty of
Agriculture, Kyoto University; Oiwakecho, Kitashirakawa, Sakyo-ku,
Kyoto, Kyoto 306-01, Japan (E-mail:misun@kais.kyoto-u.ac.jp,
Tel:075-753-6224, Fax:075-753-6226)
REFERENCE
2 (sites)
AUTHORS Shin,M., Yoshinaga,I., Uchida,A. and Ishida,Y.
TITLE Phylogenetic analysis by 16S rRNA gene sequencing of Obligate
Oligotrophs Isolated from the northern basin of Lake Biwa
(Mesotrophic Lake)
JOURNAL Unpublished (1997)
FEATURES
Location/Qualifiers
source
1..354
/organism="Spirochaeta sp."
/strain="FO-95"
/sub_species="Freshwater obligate oligotroph"
/db_xref="taxon:28185"
<1..>354
/product="16S rRNA"
BASE COUNT 91 a 79 c 111 g 73 t
ORIGIN
Query Match 92.7%; Score 20.4; DB 1; Length 354;
Best Local Similarity 95.5%; Pred. No. 25; Indels 0; Gaps 0;
Matches 21; Conservative 0; Mismatches 1;

QY 1 ttttgatcacaaagtattgata 22
|||||
Db 133 TGTGGATCACAAAGATTGATA 154

RESULT 3
AF003953 283 bp DNA BCT 30-MAY-1998
LOCUS Leptospiira interrogans strain 48/95 16S ribosomal RNA gene, partial
DEFINITION sequence.
ACCESSION AF003953
VERSION AF003953.1 GI:3169306
KEYWORDS
SOURCE Leptospiira interrogans.
ORGANISM Leptospiira interrogans;
Bacteria; Spirochaetales; Leptospiaceae; Leptospiira.
REFERENCE
1 (bases 1 to 283)
AUTHORS Woo,T.H.S., Patel,B.K.C., Cinco,M., Smythe,L.D., Symonds,M.,
Norris,M. and Dohot,M.
TITLE Identification of Leptonema by real-time homogeneous assay of rapid
cycle PCR product
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 283)
AUTHORS Woo,T.H.S., Patel,B.K.C., Cinco,M., Smythe,L.D., Symonds,M.,
Norris,M. and Dohot,M.
TITLE Direct Submission
JOURNAL Submitted (14-MAY-1997) School of Science, Griffith University,
Brisbane, QLD 4111, Australia
FEATURES
Location/Qualifiers
source
1..283
/organism="Leptospiira interrogans"
/strain="48/95"
/db_xref="taxon:173"

```


Wed Mar 28 14:03:26 2001

```

VERSION      U94977.1 GI:2735449
KEYWORDS
SOURCE
ORGANISM      Leptosira inadai.
Bacteria; Spirochaetales; Leptosiraceae; Leptosira.
REFERENCE
AUTHORS      1 (bases 1 to 288)
Woo, T.H.S., Smythe, L.D., Symonds, M., Norris, M., Dohnt, M.,
Brenner, D.J. and Patel, B.K.C.
TITLE      Identification of Leptosira inadai by continuously monitoring
fluorescence during rapid cycle PCR
JOURNAL
REFERENCE
AUTHORS      2 (bases 1 to 288)
Woo, T.H.S., Smythe, L.D., Symonds, M., Norris, M., Dohnt, M.,
Brenner, D.J. and Patel, B.K.C.
TITLE      Direct Submission
JOURNAL      Submitted (24-MAR-1997) School of Science, Griffith University,
Brisbane, QLD 4111, Australia
FEATURES
source      1..288
/organism="Leptosira inadai"
/strain="268/95"
/db_xref="taxon:29506"
<1..>288
/product="16S ribosomal RNA"
77 a 61 c 88 g 62 t

rRNA
BASE COUNT      77 a 61 c 88 g 62 t
ORIGIN
Query Match      85.5%; Score 18.8; DB 2; Length 288;
Best Local Similarity 90.9%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ttgtgatcacaaagatttgata 22
| ||||| ||||| ||||| |||||
Db 141 TATTGGATCACAGGATTGATA 162

RESULT 8
LIU94978      288 bp DNA BCT 01-JAN-1998
LOCUS      Leptosira inadai 16S ribosomal RNA gene, partial sequence.
DEFINITION
ACCESSION      U94978
VERSION
KEYWORDS
SOURCE
ORGANISM      Leptosira inadai.
Bacteria; Spirochaetales; Leptosiraceae; Leptosira.
REFERENCE
AUTHORS      1 (bases 1 to 288)
Woo, T.H.S., Smythe, L.D., Symonds, M., Norris, M., Dohnt, M.,
Brenner, D.J. and Patel, B.K.C.
TITLE      Identification of Leptosira inadai by continuously monitoring
fluorescence during rapid cycle PCR
JOURNAL
REFERENCE
AUTHORS      2 (bases 1 to 288)
Woo, T.H.S., Smythe, L.D., Symonds, M., Norris, M., Dohnt, M.,
Brenner, D.J. and Patel, B.K.C.
TITLE      Direct Submission
JOURNAL      Submitted (24-MAR-1997) School of Science, Griffith University,
Brisbane, QLD 4111, Australia
FEATURES
source      1..288
/organism="Leptosira inadai"
/strain="218/95"
/db_xref="taxon:29506"
<1..>288
/product="16S ribosomal RNA"
77 a 61 c 88 g 62 t

rRNA
BASE COUNT      77 a 61 c 88 g 62 t
ORIGIN
Query Match      85.5%; Score 18.8; DB 2; Length 288;
Best Local Similarity 90.9%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ttgtgatcacaaagatttgata 22
| ||||| ||||| ||||| |||||
Db 141 TATTGGATCACAGGATTGATA 162

RESULT 9
LIU94979      288 bp DNA BCT 01-JAN-1998
LOCUS      Leptosira inadai 16S ribosomal RNA gene, partial sequence.
DEFINITION
ACCESSION      U94979
VERSION
KEYWORDS
SOURCE
ORGANISM      Leptosira inadai.
Bacteria; Spirochaetales; Leptosiraceae; Leptosira.
REFERENCE
AUTHORS      1 (bases 1 to 288)
Woo, T.H.S., Smythe, L.D., Symonds, M., Norris, M., Dohnt, M.,
Brenner, D.J. and Patel, B.K.C.
TITLE      Identification of Leptosira inadai by continuously monitoring
fluorescence during rapid cycle PCR
JOURNAL
REFERENCE
AUTHORS      2 (bases 1 to 288)
Woo, T.H.S., Smythe, L.D., Symonds, M., Norris, M., Dohnt, M.,
Brenner, D.J. and Patel, B.K.C.
TITLE      Direct Submission
JOURNAL      Submitted (24-MAR-1997) School of Science, Griffith University,
Brisbane, QLD 4111, Australia
FEATURES
source      1..288
/organism="Leptosira inadai"
/strain="1078 VPI"
/db_xref="taxon:29506"
<1..>288
/product="16S ribosomal RNA"
77 a 61 c 88 g 62 t

rRNA
BASE COUNT      77 a 61 c 88 g 62 t
ORIGIN
Query Match      85.5%; Score 18.8; DB 2; Length 288;
Best Local Similarity 90.9%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ttgtgatcacaaagatttgata 22
| ||||| ||||| ||||| |||||
Db 141 TATTGGATCACAGGATTGATA 162

RESULT 10
AB007012      353 bp DNA BCT 13-OCT-1997
LOCUS      Spirochaeta sp. 16S rRNA gene, partial sequence.
DEFINITION
ACCESSION      AB007012
VERSION
KEYWORDS
SOURCE
ORGANISM      Spirochaeta sp.
Bacteria; Spirochaetales; Spirochaetaceae; Spirochaeta.
REFERENCE
AUTHORS      1 (bases 1 to 353)
Shin, M.-S.
TITLE      Direct Submission
JOURNAL      Submitted (05-SEP-1997) to the DDBJ/EMBL/GenBank databases. Mi-Sun
Shin, Laboratory of Marine Molecular Microbiology, Faculty of
Agriculture, Kyoto University, Oiwakecho, Kitashirakawa, Sakyo-ku,
Kyoto, Kyoto 606-01, Japan (E-mail: misun@kais.kyoto-u.ac.jp,
Tel: 075-753-6224, Fax: 075-753-6226)
2 (sites)
Shin, M., Yoshinaga, I., Uchida, A. and Ishida, Y.
Phylogenetic analysis by 16S rRNA gene sequencing of Obligate
Oligotrophs isolated from the northern basin of Lake Biwa
(Mesotrophic Lake)
Unpublished (1997)

```

```

FEATURES
  source      Location/Qualifiers
 1..353
   /organism="Spirochaeta sp."
   /strain="SO-104"
   /sub_species="Freshwater obligate oligotroph"
   /db_xref="taxon:28185"
   <L..>353
   /product="16S rRNA"
 90 a      80 c      108 g      75 t
ORIGIN
Query Match      85.5%; Score 18.8; DB 1; Length 353;
Best Local Similarity 90.9%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 tgttgatcacagattgata 22
||||| ||||| ||||| |||||
Db 132 TGTGTGATCACAGATTCGATA 153

RESULT 11
ATHRS2X
LOCUS      ATHRS2X      1012 bp      mRNA      PLN      30-OCT-1994
DEFINITION Arabidopsis thaliana ribonuclease (RNS2) mRNA, complete cds.
ACCESSION M98336
VERSION M98336.1 GI:289209
KEYWORDS  ribonuclease.
SOURCE      Arabidopsis thaliana cDNA to mRNA.
ORGANISM    Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 1012)
AUTHORS    Taylor,C.B., Bariola,P.A., delCardayre,S.B., Raines,R.T. and
Green,P.J.
TITLE      RNS2: a senescence-associated RNase of Arabidopsis that diverged
from the S-RNases before speciation
JOURNAL    Proc. Natl. Acad. Sci. U.S.A. 90 (11), 5118-5122 (1993)
MEDLINE    93281708
FEATURES
  source      Location/Qualifiers
 1..1012
   /organism="Arabidopsis thaliana"
   /db_xref="taxon:3702"
 16..795
   /gene="RNS2"
 16..795
   /gene="RNS2"
/codon_start=1
/product="ribonuclease"
/protein_id="AAA51406.1"
/db_xref="GI:289210"
/translation="NASRLCLLLVACIAGFAGDVIELNRSOREFDYFALSLOWPQT
YCRGTRHCCSKNACCGSDAPQTQTHGLWPDYNDGSMPCCYRSDRDEKEKETSLMDG
LEKYWPSLCGSPSCNGKGSFGEHWEKHGTCSPVPHDEYNVFLTLNLKHNV
TDVYOAGTVASNSKEYPLGGIVTAIQNAFHITPEVVCVKRDAIDIRICFCYKDFKPRD
CVGSDQLTSRKSCPIYVSLPEYTPLDGAMVLKMPTERAL"
287 a      197 c      217 g      311 t
ORIGIN
Query Match      85.5%; Score 18.8; DB 7; Length 1012;
Best Local Similarity 90.9%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 tgttgatcacagattgata 22
||||| ||||| ||||| |||||
Db 672 TGTGTGATCACAGATTCGATA 693

RESULT 12
AC003000
LOCUS      AC003000      72590 bp      DNA      PLN      05-APR-2000

```

```

DEFINITION
Arabidopsis thaliana chromosome II section 214 of 255 of the
complete sequence. Sequence from clones T517.
ACCESSION
AC003000 AE002093
VERSION
AC003000.2 GI:6598383
KEYWORDS
HTG
SOURCE
thale cress.
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 72590)
AUTHORS    Lin,X., Kaul,S., Rounsley,S.D., Shea,T.P., Benito,M.-I., Town,C.D.,
Fujii,C.Y., Mason,T.M., Bowman,C.L., Barnstead,M.E.,
Feldblum,T.V., Buell,C.R., Ketchum,K.A., Lee,J.J., Ronning,C.M.,
Koo,H., Moffat,K.S., Cronin,L.A., Shen,M., VanAken,S.E., Umayam,L.,
Tallon,L.J., Gill,J.E., Adams,M.D., Carrera,A.J., Cressy,T.H.,
Goodman,H.M., Somerville,C.R., Copenhaver,G.P., Preuss,D.,
Nierman,W.C., White,O., Eisen,J.A., Salzberg,S.L., Fraser,C.M. and
Venter,J.C.
TITLE      Sequence and analysis of chromosome 2 of the plant Arabidopsis
thaliana
JOURNAL    Nature 402 (6763), 761-768 (1999)
MEDLINE    20083487
PUBMED     10617197
REFERENCE 2 (bases 1 to 72590)
AUTHORS    Lin,X.
JOURNAL    Direct Submission
COMMENT    Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
On Dec 17, 1999 this sequence version replaced gi:2642152.
The sequence and annotation of this chromosome 2 were merged from those
of the individual clones on this chromosome after removing
overlaps. For detailed information, please see the TIGR web site
(http://www.tigr.org/tldb/at/at.html).

Genes were identified by a combination of three methods: Gene
prediction programs including GRAIL
(http://arthur.epm.ornl.gov/pub/xgrail), Genefinder (Phil Green,
University of Washington), Genscan (Chris Burge,
http://gnomic.stanford.edu/GENSCANW.html), and NetPlantGene
(http://www.cbs.dtu.dk/services/NetGene2/), searches of the
complete sequence against a peptide database and plant EST
databases at TIGR, and manual curations based on those analyses.
Annotated genes are named to indicate the level of evidence for
their annotation. Genes with similarity to other proteins are named
after the database hits. Genes without significant peptide
similarity but with EST similarity are named as 'unknown' proteins.
Genes without protein or EST similarity, that are predicted by two
or more gene prediction programs over most of their length are
annotated as 'hypothetical' proteins. Genes encoding tRNAs are
predicted by tRNAscan-SE (Sean Eddy,
http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats were
identified by repeatmasker (Arian Smit,
http://ftp.genome.washington.edu/RM/RepeatMasker.html). Genes are
numbered from the top to bottom of the chromosome.

We thank the CSHL/WashU/ABI consortium for sequencing BAC clones
F6P23, F5J6, T17A5, and T13L16, the ESSA group for sequencing clone
F13D4, and Scott Jackson, Jiming Jiang, Klaus Meyer, Eric Richards
and Satoshi Tabata for helpful assistance. In addition, we would
like to thank the TIGR Bioinformatics Department, especially Lixin
Zhou, Hanif Khalak, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy
Peterson, Michael Holmes, and Delwood Richardson for software and
database support.

This work was supported by the National Science Foundation,
Department of Energy and the US Department of Agriculture.

Address all correspondence to: at@tigr.org.
Location/Qualifiers
1..72590
/organism="Arabidopsis thaliana"
/cultivar="Columbia"

```

```

/db_xref="taxon:3702"
/chromosome="II"
/notes="Sequence from clone T517"
complement(130..198)
/rpt_family="(CAAA)n"
complement(1006..1024)
/rpt_family="POLY_A"
complement(1266..1291)
/rpt_family="(CAAAA)n"
complement(2759..2862)
/rpt_family="POLY_A"
complement(4584..>5909)
/genes="At2g39710"
complement(4584..>5909)
/genes="At2g39710"
complement(4584..>5909)
/notes="T517.1; similar to
GP12245012[gnl|PID|e327000|297341"
complement(4584..5909)
/genes="At2g39710"
/notes="unknown protein"
/codon_start=1
/protein_id="AAB87120.1"
/db_xref="GI:2642153"
/translacion="MDSSSSSLSKSNFLRISVLLIPLFLFCKTSSTNQTLFSLK
TOKLQSSSSDKLFRINUTLVAVGDDPPQNTSMVLDTCSELSLHCKSSNGLSVF
NPVSSVSVPCSSPTRTRDLPAPACDPKTHLCHVAISYADATKTEGNLAHET
FVIGSVTRPGLTGCMDSGILSSNEEDAKSTGLMGNRGLSLFVNOLGFSKYSLSLPG
SSSGFLLDGASTSLGPIQYPLVLOSTPLPYFORVAYTVOLEGRVGSKLISLCP
SVFPHDTGAGQTMVDSGTQFTLMLGPVYALKNETOTKSVLRVLDVDDFVQCTM
DLCKVGSTTRPNFSPMLPMSLPRGAEVSGOKLIRYVAGSGSGKBEVYCTGNG
SDLGIGFAVFIHHQGNVMEFLAKRSYGFAGNVKRCDLASRLGLRD"
complement(<10039..11304)
/genes="At2g39720"
complement(<10039..11304)
/genes="T517.2"
complement(10039..11244)
/genes="At2g39720"
/notes="unknown protein"
/codon_start=1
/protein_id="AAB87121.1"
/db_xref="GI:2642154"
/translacion="MASGSCWYCSRFVWVSDISCPDCDGGLELIQRLDPTPSD
SPTATTTTHQSPTRFPFPPSSSTPSASMHADNSPTTIVTRRSRNSPNPVILRG
SAAATPSQSVSEGLDRFAPQYDDGTDSGLRPLPPSEITLGGDFDLDLQSLQE
LNTNLNLSCHPAPKSAIIBALPIELDPTHLSDSQCACVCKNFVLKSSAREMP
CNHIIYPCILPWLAIKNSCPVRLPAEDLTDGTAALFVATATAEEDSSAAGLT
INRLPGGFGFVGRIPGWRGDRMPPVYTVEDGRLDERLPRVWAGSRGRGDGG
GSRRRNWLA"
complement(11453..11512)
/rpt_family="POLY_A"
complement(13323..13599,13724..14005,14087..14559,
14651..14736,14836..14941,15033..15349,15834..16055))
/genes="At2g39730"
complement(13323..13599,13713..14005,14087..14559,
14651..14736,14836..14941,15033..15349,15834..16055))
/genes="At2g39730"
complement(13323..16055)
/genes="At2g39730"
/notes="T517.3"
13364..13461
/rpt_family="POLY_A"
complement(13484..13599,13724..14005,14087..14559,
14651..14736,14836..14941,15033..15349,15834..15878))
/genes="At2g39730"
/notes="alternative splice form 1 GP166834|166720"
/codon_start=1
/product="Rubisco activase"
/protein_id="AAB87122.1"
/db_xref="GI:2642155"

```



```

/ gene="ftsI"
1733 .3862
/ gene="ftsI"
/ codon_start=1
/ product="putative peptidoglycan synthetase"
/ protein_id="AAD54780.1"
/ db_xref="GI:5880687"
/ translation="MIYNPMKYRRFRRLPKHLQIYYAFLSISTMIKIALDPYSKR
PMKWHSGTPOYENBERMVKLSLATVGLLFATRLSGQFNKYTELSPKLSVAERQIQG
VNDQQRKILIDHGDIVADLPAYDLYVHPRMCSISLERIAELLSPILDSQOYLIN
RLDGGSCICLMDHIDINTSAQIPRLGVDCIELVHHPQVYPRKGSFESTLGVVDTEG
YQAGLESSLDWMSKSYQDPCWMDGHGFLGRPKQILFHQESALQITIDCGLOE
KVSOLITNMRNFKAKRIALIMEAGSLRCLATISFSDPNCYGPMPFERFCWPT
DLFFPGSTFKFVNIAIALENGIFOPTDRILDTKIRGDSWIGVNGGFWIWRSLBHL
TKDITSDYATISVVKDQDQFVDHIEAATFSGQLAMTLPKLQLIATLANGMAVT
PHLISKVITLDFHHLQSMNEFSIQGWGQSVLSRQYHAKQPRPYTHDLYLGHVPVP
SLELGFVFKSTIPHTERRRLFSRQTCNVLLGMLQVLDQAQATSGRGLPGYAMAG
KGTAKASALGGSTDSVTSFVGIYPAVKPKFTVLIIDEDEDPREFGENTAVDVT
QTLISEMIVQEDPPSPYTSLSLFERN"
3998 .5059
/ gene="psbA"
3998 .5059
/ gene="psbA"
/ codon_start=1
/ product="D1 reaction-center protein of photosystem II"
/ protein_id="AAD54781.1"
/ db_xref="GI:5880688"
/ translation="MTAILERRESTVWARFCDWVTSTENKLYIGWFCVLMPILLITA
TSVFIIGFAAPVDIDIREPVSGLIFGNNIISGAILPSSAAIGIHFIPIWEAASI
DEWLNGGCVELIVHFLGVACVYMGREWLSFRGLMRPWAVAYSAPVAAATAVFLI
YFGQSGSDGMLPGISGTFNFMIVFOAEHNLHPHFMGLVAGVFGGSLFSAMHSL
VTSILRETTENESANAGYKFCQBEETNIVAACHYFGRLLIFQVASFNNSSLSHFFLA
AMPVCVIFETALGVSTMAFNLGFCNFSVSDSQSRVINTWADIINRANLMEVYHER
NAHPFDLIASVDAPAVOG"
5489 .6964
/ gene="atpB"
5489 .6964
/ gene="atpB"
/ codon_start=1
/ product="CF1 beta subunit of ATP synthase"
/ protein_id="AAD54782.1"
/ db_xref="GI:5880689"
/ translation="MSQAGVKTKQAGKIVQIVGPMVDVAFQPGSMPTNVALIVKGRN
GAGQEVSVNCEVOQLLGDGLVRAVMSATDGLMRGEVDTDRALSPVPGPITLGRIG
NAGLEVPDNGVGNKTLPIHREAPAFVLDLTKLSTFETGKVVVDLLAPYRGGKADY
LFCGAGVKTVLIMELNNTAKAHGCVSVFGVGGERTREGNDLYMCECKSVINKADY
KSSKVALYQOMNEPPGAMRVCGLTALTAERYFRDYNODVLLFIDNIERFVQAGSEV
SALLGRMPSAVGIQPLATENGLOERITSTKDGSLTSIOAVYVPADDLTDPAPATTF
AHLDTATVLSRLAAKGIYPAPVDPLDSTMLQPLWILGDDHYGTAQRVKQTLQRYKEL
QDITAILGLDELSEEDRMIVARARKIERFLSPQFFVAEFTGAPGKYVPLAESIQGFK
LILSGELDSLPESAFLVGNIEPAIAKASIAQAAAAAAA"
6977 .7378
/ gene="atpE"
6977 .7378
/ gene="atpE"
/ codon_start=1
/ product="CF1 epsilon subunit of ATP synthase"
/ protein_id="AAD54783.1"
/ db_xref="GI:5880690"
/ translation="MSLQVRILTPEQVFLNVSADAILPTNTGMGVLTNHTPLITAI
DIGPMLVRSESTWQSMALLGGLALVKDNOVILVNEAELGSDINAEAEITFLAAKEA
LANSKRDKQIENNLAKFRKRVYQVATLVK"
complement(7622 .8104)
/ gene="petB"
complement(7622 .8104)
/ gene="petB"
/ codon_start=1
/ product="subunit IV of cytochrome b6/f complex"
/ protein_id="AAD54784.1"
/ db_xref="GI:5880691"
/ translation="MSVTKKPDLTDVPLVRLAKLAKGMGHNYGEPANWDLNLYMFPVVI
LGTLSCTITGLAVLDPAAIGEPANFPATPLILPEWYFPFQLLRITVPNKLGLVLLMA
AVPAGLLTVPFIESINKFQNPFRPVATTVELICTVVAIWLIGIGATLPIDISLTGLF
complement(8132 .8779)
/ gene="petB"
complement(8132 .8779)
/ gene="petB"
/ codon_start=1
/ product="apocytochrome b6 of cytochrome b6/f complex"
/ protein_id="AAD54785.1"
/ db_xref="GI:5880692"
/ translation="MSKIYDNFEERLEIQATADDITSKYVPPHVNIFVCFGIGITLTCF
LIQVATGAMFYIYRPTVEAFSVEIYTNVNWGNLIRSIRHRSASMMVMMLLHVFE
RYLVTGGFKPRELTVTGTAVILAVITVSFGVSLPMDQVGYWAKYIVTGVPAIPV
IGAPVLEILRSVSVSQSTLFRVSLHTFVLPLLTAVFMLHFLMIRKQGISGPL"
complement(9045 .9431)
/ gene="rpl12"
complement(9045 .9431)
/ gene="rpl12"
/ codon_start=1
/ product="ribosomal protein L12"
/ protein_id="AAD54786.1"
/ db_xref="GI:5880693"
/ translation="NSTITNQIIEQKSMTLLEAAELVSIETFTGVDSAPAGAVM
VRAAPSPVVEOTEFTVINDVPSAKRIAVIKVRALTLGLKREAKDMIESVPKAI
KEGIGKDEAQIKQOLEEAGATVTK"
complement(9442 .9846)
/ gene="rps9"
complement(9442 .9846)
/ gene="rps9"
/ codon_start=1
/ product="ribosomal protein S9"
/ protein_id="AAD54787.1"
/ db_xref="GI:5880694"
/ translation="MDHTSSVGLRKTGRKCAVAOYRIIECTCOLTINORPGVPYLOV
NPAYLLAAQALDVTETLNTYDITVKTGGGLTGQAEALNGLARALCSLHKHKKAL
KPYGLTRNPLRKRKKYGLKARKAPQFSKR"
complement(9998 .11485)
/ gene="rpoA"
complement(9998 .11485)
/ gene="rpoA"
/ codon_start=1
/ product="alpha subunit of RNA polymerase"
/ protein_id="AAD54788.1"
Query Match 85.5%; Score 18.8; DB 6; Length 200799;
Best Local Similarity 90.9%; Pred. No. 77;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 tggttgatcacagaattgata 22
||||| ||||| ||||| |||||
Db 134224 TGTGGCTCATGATTGATA 134245
RESULT 15
AF137379 200799 bp DNA circular PLN 14-SEP-1999
LOCUS Nephroselmis olivacea chloroplast DNA, complete genome.
DEFINITION AF137379
ACCESSION AF137379
VERSION AF137379.1 GI:5880684
KEYWORDS Nephroselmis olivacea.
SOURCE Chloroplast Nephroselmis olivacea
ORGANISM Eukaryota; Viridiplantae; Chlorophyta; Prasinophyceae;
Chlorodendreales; Chlorodendraceae; Nephroselmis.
REFERENCE 1 (bases 1 to 200799)
AUTHORS Turmel, M., Otis, C. and Lemieux, C.
TITLE The complete chloroplast DNA sequence of the green alga
Nephroselmis olivacea: insights into the architecture of ancestral
chloroplast genomes
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (18), 10248-10253 (1999)
MEDLINE 99398694
REFERENCE 2 (bases 1 to 200799)
AUTHORS Turmel, M., Otis, C. and Lemieux, C.
```

TITLE Direct Submission
 JOURNAL Submitted (24-MAR-1999) Biochimie, Universite Laval, Pavillon
 Charles-Eugene Marchand, Quebec G1K 7P4, Canada
 FEATURES
 Location/Qualifiers
 1..200799
 /organism="Nephroselmis olivacea"
 /organelle="plastid:chloroplast"
 /db_xref="taxon:31312"
 /cell_line="NIES 484"
 misc_feature
 1..92126
 /note="Region: large single-copy region (LSC)"
 tRNA
 55..137
 /gene="trnL(gag)"
 /note="codons recognized: CUY"
 /product="tRNA-Leu"
 /anticodon=(pos:88..90,aa:Leu)
 gene
 55..137
 /gene="trnL(gag)"
 332..634
 /gene="rps14"
 332..634
 /gene="rps14"
 /codon_start=1
 /product="ribosomal protein S14"
 /protein_id="AAD54778.1"
 /db_xref="GI:5880685"
 /translation="WAKKSMIERESKRAALVAKYATRRQALKAAIOKTKSFDERLVLO
 HOQLDPVNSVPCRLHNRCTITGRPGKYRDFGLSRHELRAHAGCLLPVTKASW"
 tRNA
 765..849
 /gene="trnL(uaa)"
 /note="codon recognized: UUA"
 /product="tRNA-Leu"
 /anticodon=(pos:798..800,aa:Leu)
 gene
 765..849
 /gene="trnL(uaa)"
 tRNA
 983..1054
 /gene="trnQ(uug)"
 /note="codons recognized: CAR"
 /product="tRNA-Gln"
 /anticodon=(pos:1015..1017,aa:Gln)
 gene
 983..1054
 /gene="trnQ(uug)"
 tRNA
 1102..1182
 /gene="trnL(uag)"
 /note="codons recognized: CUR"
 /product="tRNA-Leu"
 /anticodon=(pos:1136..1138,aa:Leu)
 gene
 1102..1182
 /gene="trnL(uag)"
 complement(1364..1468)
 /gene="psbM"
 complement(1364..1468)
 /gene="psbM"
 /codon_start=1
 /product="M protein of photosystem II"
 /protein_id="AAD54779.1"
 /db_xref="GI:5880686"
 /translation="MEVNILGLIATLFIITPSFLILYVKTASQON"
 tRNA
 1618..1691
 /gene="trnM(cau)"
 /note="codon recognized: AUG"
 /product="tRNA-Met"
 /anticodon=(pos:1652..1654,aa:Met)
 gene
 1618..1691
 /gene="trnM(cau)"
 gene
 1733..3862
 /gene="ftsI"
 1733..3862
 /gene="ftsI"
 /codon_start=1
 /product="putative peptidoglycan synthetase"
 /protein_id="AAD54780.1"
 /db_xref="GI:5880687"
 /translation="MIYNPMKYRQFRRLPKHLOGIYYAFISISITMIXIALDPYSKR
 PKMWHSGTPFOYENERMVMIKLSLATVGLLFATRLSGLOFNKYTELKSAVERQQICQ
 VNDPQOKIILDHGDIIVADILPAYDLYVHPRMCSISLERIAELISPIDLSQOYLIN
 RLDGSGICLMHQIDNTSAQITRRLGVDGIELVHPQRYVPKRGSFESILGVDTGE
 YGAGLESSLDDMMKSTYODPCWMDGHGNGFIRPKOILFHOESALQLTIDCGLOE
 KYSQILTNAMNREGAKRIAAIINEAHSGAIRCLATSPSYDPCYGFWPFRCWPIT
 DTQFQSTFKPNALALENGIFQPTDRILDITKIRIGDSWICNGVGGGFITWRSLDHL
 TCTQILQRSSNVMVRMOSLDPAIYHRNLIRIGLSHRNDNQTSFKMSHDHNSGW
 NLKDLTSDYAISVVKDQDEFVDIEAATAFSGOGLAMTPLKLLQIATITANGMAVT
 PHILISKIVTDHFHLOSNEFSLOGWQGSVLSRQYHAKOPRPRTYHDLYLGHVVP
 SLELGFWDVKSIPPHTRRRLRFSQTCNVLLGMLQOVLDAAQATSGRFLPCYANAG
 KTGTAQKASALGGYSTDSVTSEVGLIYPAVKPFVTLVLIIDPEDPFRFGNTAVDVT
 QTLISEMIVQEQDPPSYPTVSLFRNM"
 gene
 3998..5059
 /gene="psbA"
 3998..5059
 /gene="psbA"
 /codon_start=1
 /product="D1 reaction-center protein of photosystem II"
 /protein_id="AAD54781.1"
 /db_xref="GI:5880688"
 /translation="MTAILERRESTSVARFCDWVTSTENRLYIGWFGVLMIPLLLTA
 TSVFIIGFIAAPPVDIDGIREPVSGSLFGNNIISGAIIPSSAAIGIHFPYIWEAASI
 DEMLYNGCYELLIVLHLLGVACVYCREHELFSRLCMRPHIAYASAPAAATAVELI
 YPIQGSGSDGMPLGISGTFNEMIVFQACHNIIHPFRLMGVAGVFGSLFSAMRSL
 VTSLIRETTENESANAGYFGQEEETYNIVAAHGYFGRLLIFOYAFNNSRSLHFFLA
 AMPVVICWFTALGVSTMAFNLGFNFQNSQVDSQGRVINTWADIINRANLGMVMEH
 NAHNPLDLASVDAVAVQ"
 gene
 5489..6964
 /gene="atpB"
 5489..6964
 /gene="atpB"
 /codon_start=1
 /product="CF1 beta subunit of ATP synthase"
 /protein_id="AAD54782.1"
 /db_xref="GI:5880689"
 /translation="MSQAGVKTKQAKIVQIVGPMVDVAFQPGSMNIYNALIVKGRN
 GAGVEVSVCEVQQLIGDLGVLRASMSATDGLMRGMEVTDTRALSVPGPTLGRIF
 NIVGEPVDMNPGVNEKTLPIHREAPFVLDLTKLIFETGKVLDLLAPYRRGGIG
 LFGAGVGKTVLIMELINIAKAGHGVSGVGGVTRREGNDLYMCECSKINKADY
 KSKVALVYQMNPPGARMVGLTALMAEYFRDYNNDVLLFIQNIREFVQAGSEV
 SALLGRMPSAVGYOPTLATFMGLOERITSTKDGSI'ISOAVYPADDLTDPAATFF
 AHLDATTVLSRLAAKGIYPAVDPLDSTMTQPLWILGDDHYGTAQRVKTQRYKEL
 QDITAILGLDELSEEDRMIVARARKIERFSLQPFVFAEVTGAPGYVPLAESIQGFK
 LILSGELSLPESAFYLVGNIETAKAASIAQAAAAA"
 gene
 6977..7378
 /gene="atpE"
 6977..7378
 /gene="atpE"
 /codon_start=1
 /product="CF1 epsilon subunit of ATP synthase"
 /protein_id="AAD54783.1"
 /db_xref="GI:5880690"
 /translation="MSLOVRILTPEOVFLNVSADETILPTNGCMGLTNHTPLTIAI
 DIGPMLVRSESTWQSMALLGLYALKDQNOVILVNEAELGSDINAEAEETFLAEEA
 LANSKTRKQDIENNLAFAKRVYQATLVK"
 gene
 complement(7622..8104)
 /gene="petD"
 complement(7622..8104)
 /gene="petD"
 /codon_start=1
 /product="subunit IV of cytochrome b6/f complex"
 /protein_id="AAD54784.1"
 /db_xref="GI:5880691"
 /translation="MSVTRKKPDLTDPVLRAKLAKGMHNVYGEPAWPNDLLYMPVVI
 LGLTSCITGLAVLDPAALIGEPAWPAATPLEIPWYFFPVFOLLTPVKNKLLGVLLMA
 AVPAGLLTVPFIESINKFQNPFRPVATTVFLIGTVVAIWLGIGATLPIDLSLTGLF"
 complement(8132..8779)
 /gene="petB"
 complement(8132..8779)
 /gene="petB"
 /codon_start=1

```
/product="apocytochrome b6 of cytochrome b6/f complex"
/protein_id="AAD54785.1"
/db_xref="GI:5880692"
/translation="MSKIYDFEERLEIQAIADDTISKYVPPHVNIEYCEGGITLTCF
LIQVATGFMTFYRTAFASVEYIMTNVFGWLIRSIHRWSASMMVMMLILLHF
RVLTGGFKRPRLTWTVTGVLAVITVSGVTGSLPDQGVYAVKIVTVGVDAIPV
IGAPLVLLRGSVSGSTLTRYSLHTFVPLLTAVFMLMHLMIKQGISGPL"
complement(9045..9431)
/gene="rpl12"
complement(9045..9431)
/gene="rpl12"
/codon_start=1
/product="ribosomal protein L12"
/protein_id="AAD54786.1"
/db_xref="GI:5880693"
/translation="MSITNQIIEQLKSMILLAAELVSOIETFGVDASAPAGAVM
VAAAPSPAVVEEQTEFTVIMINDVPSAKRIAVIKVRALIGLGLKEAKDMIESVPKAI
REGIGKDAQOIKOOLEEAGATTVK"
complement(9442..9846)
/gene="rps9"
complement(9442..9846)
/gene="rps9"
/codon_start=1
/product="ribosomal protein S9"
/protein_id="AAD54787.1"
/db_xref="GI:5880694"
/translation="MDHTSSVGLRKTGRKCAVAQVRIEGTGOLTINQREGVPYLYQ
NPAYLLAAGALDVTETLNRYDTIVKTKGGTLTGCAEAIKGLARALCSLHKRKAL
KPGYLTRNPLRKRRKKYGLKARKAPQFSKR"
complement(9998..11485)
/gene="rpoA"
complement(9998..11485)
/gene="rpoA"
/codon_start=1
/product="alpha subunit of RNA polymerase"
/protein_id="AAD54788.1"
```

Query Match 85.5% Score 18.8; DB 6; Length 200799;
Best Local Similarity 90.9%; Pred. No. 77;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 tgttgatcacagaatttgata 22
||||| ||||| ||||| |||||
Db 158702 TGTGGCTCACATGATTGATA 158681

Search completed: March 27, 2001, 08:17:15
Job time: 5912 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 27, 2001, 08:17:04 ; Search time 2286.42 Seconds
(without alignments)
15.668 Million cell updates/sec

Title: US-09-380-826A-5
Perfect score: 7
Sequence: 1 ttgata 7

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1118133 seqs, 2558875100 residues

Total number of hits satisfying chosen parameters: 2236266

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_bal:*
- 2: gb_baz:*
- 3: gb_om:*
- 4: gb_ov:*
- 5: gb_ph:*
- 6: gb_pl1:*
- 7: gb_pl2:*
- 8: gb_pl3:*
- 9: gb_pr1:*
- 10: gb_pr2:*
- 11: gb_pr3:*
- 12: gb_ro:*
- 13: gb_sy:*
- 14: gb_un:*
- 15: em_fun:*
- 16: em_hum1:*
- 17: em_hum2:*
- 18: em_in:*
- 19: em_om:*
- 20: em_or:*
- 21: em_ov:*
- 22: em_pat:*
- 23: em_ph:*
- 24: em_pi:*
- 25: em_ro:*
- 26: em_sts:*
- 27: em_sy:*
- 28: em_un:*
- 29: gb_ba3:*
- 30: gb_in1:*
- 31: gb_in2:*
- 32: gb_in3:*
- 33: gb_pl3:*
- 34: gb_pr4:*
- 35: em_bal:*
- 36: em_ba2:*
- 37: em_htg1:*
- 38: em_htg2:*
- 39: em_htg3:*
- 40: em_htg4:*
- 41: em_htg5:*
- 42: em_htg6:*
- 43: em_htg7:*

- 44: em_htg8:*
- 45: em_htg9:*
- 46: em_htg10:*
- 47: em_hum3:*
- 48: em_hum4:*
- 49: em_hum5:*
- 50: em_hum6:*
- 51: gb_pr5:*
- 52: gb_pr6:*
- 53: gb_pr7:*
- 54: gb_htg1:*
- 55: gb_htg2:*
- 56: gb_htg3:*
- 57: gb_htg4:*
- 58: gb_htg5:*
- 59: gb_htg6:*
- 60: gb_htg7:*
- 61: gb_htg8:*
- 62: gb_htg9:*
- 63: gb_htg10:*
- 64: gb_htg11:*
- 65: gb_htg12:*
- 66: gb_htg13:*
- 67: gb_htg14:*
- 68: gb_htg15:*
- 69: gb_htg16:*
- 70: gb_htg17:*
- 71: gb_htg18:*
- 72: gb_htg19:*
- 73: gb_htg20:*
- 74: gb_htg21:*
- 75: gb_htg22:*
- 76: gb_htg23:*
- 77: gb_sts1:*
- 78: gb_sts2:*
- 79: gb_vil:*
- 80: gb_vil2:*
- 81: gb_pat1:*
- 82: gb_pat2:*
- 83: em_htg0:*
- 84: gb_htg24:*
- 85: gb_pr8:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
C 1	7	100.0	11	81 I20741	I20741 Sequence 6
C 2	7	100.0	11	81 I20742	I20742 Sequence 7
C 3	7	100.0	13	81 AR082373	AR082373 Sequence 7
C 4	7	100.0	13	81 I52063	I52063 Sequence 5
C 5	7	100.0	13	82 I78419	I78419 Sequence 21
C 6	7	100.0	14	79 AX007639	AX007639 Sequence 74
C 7	7	100.0	14	81 A88596	A88596 Sequence 74
C 8	7	100.0	14	81 A90563	A90563 Sequence 74
C 9	7	100.0	14	81 AR082368	AR082368 Sequence 21
C 10	7	100.0	14	82 I78414	I78414 Sequence 21
C 11	7	100.0	15	79 AX007638	AX007638 Sequence 12
C 12	7	100.0	15	79 AX007675	AX007675 Sequence 13
C 13	7	100.0	15	81 A35660	A35660 Synthetic h
C 14	7	100.0	15	81 AR045297	AR045297 Sequence 90
C 15	7	100.0	15	81 AR071525	AR071525 Sequence 13
C 16	7	100.0	15	81 I52349	I52349 Sequence 13
C 17	7	100.0	15	81 I61575	I61575 Sequence 13
C 18	7	100.0	15	81 I61576	I61576 Sequence 13
C 19	7	100.0	15	81 I61577	I61577 Sequence 8
C 20	7	100.0	15	82 I77301	I77301 Sequence 9
C 21	7	100.0	15	82 I77302	I77302 Sequence 9

[illegible]

QY 1 ttgtata 7
 |||||
 Db 10 TTTGATA 4

RESULT 5
 I78419/c 178419 13 bp DNA PAT 03-APR-1998
 LOCUS Sequence 217 from patent US 5693535.
 DEFINITION I78419
 ACCESSION I78419
 VERSION I78419.1 GI:3014573
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unclassified.
 REFERENCE 1 (bases 1 to 13)
 AUTHORS Draper,K.G., Chowrira,B., McSwiggen,J., Stinchcomb,D.T. and Thompson,J.D.
 TITLE HIV targeted ribozymes
 JOURNAL Patent: US 5693535-A 217 02-DEC-1997;
 FEATURES Location/Qualifiers
 source 1..13
 /organism="unknown"
 BASE COUNT 5 a 4 c 1 g 3 t
 ORIGIN

Query Match 100.0%; Score 7; DB 82; Length 13;
 Best Local Similarity 100.0%; Pred. No. 6.2e+05;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtata 7
 |||||
 Db 10 TTTGATA 4

RESULT 6
 AX007639/c AX007639 14 bp DNA VRL 06-SEP-2000
 LOCUS Sequence 181 from Patent WO9967428.
 DEFINITION AX007639
 ACCESSION AX007639
 VERSION AX007639.1 GI:9995336
 KEYWORDS Aids-associated retrovirus.
 SOURCE Aids-associated retrovirus.
 ORGANISM Viruses; Retroid viruses; Retroviridae.
 REFERENCE 1 (bases 1 to 14)
 AUTHORS Stuyver,L.
 TITLE Method for detection of drug-selected mutations in the hiv protease gene
 JOURNAL Patent: WO 9967428-A 29-DEC-1999;
 INNOCENETICS NV (BE); STUYVER LIEVEN (BE)
 FEATURES Location/Qualifiers
 source 1..14
 /organism="Aids-associated retrovirus"
 /db_xref="taxon:11966"
 BASE COUNT 5 a 1 c 2 g 6 t
 ORIGIN

Query Match 100.0%; Score 7; DB 79; Length 14;
 Best Local Similarity 100.0%; Pred. No. 6.2e+05;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtata 7
 |||||
 Db 11 TTTGATA 5

RESULT 7
 A88596 14 bp DNA PAT 22-JAN-2000
 LOCUS Sequence 212 from patent US 5972704.
 DEFINITION A88596
 ACCESSION A88596
 VERSION A88596.1 GI:10009094
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unclassified.
 REFERENCE 1 (bases 1 to 14)
 AUTHORS Draper,K.G., Chowrira,B., McSwiggen,J., Stinchcomb,D.T. and

DEFINITION Sequence 744 from Patent WO9833904.
 ACCESSION A88596
 VERSION A88596.1 GI:6737166
 KEYWORDS
 SOURCE unidentified.
 ORGANISM unidentified.
 REFERENCE 1 (bases 1 to 14)
 AUTHORS Brysch,W. and Schlingsenslepen,K.
 TITLE AN ANTISENSE OLIGONUCLEOTIDE PREPARATION METHOD
 JOURNAL Patent: WO 9833904-A 06-AUG-1998;
 BIOGNOSTIK GES (DE); BRYSCH WOLFGANG (DE)
 FEATURES Location/Qualifiers
 source 1..14
 /organism="unidentified"
 /db_xref="taxon:32644"
 BASE COUNT 5 a 1 c 1 g 7 t
 ORIGIN

Query Match 100.0%; Score 7; DB 81; Length 14;
 Best Local Similarity 100.0%; Pred. No. 6.2e+05;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtata 7
 |||||
 Db 1 TTTGATA 7

RESULT 8
 A90563 A90563 14 bp DNA PAT 22-JAN-2000
 LOCUS Sequence 744 from Patent EP0856579.
 DEFINITION A90563
 ACCESSION A90563
 VERSION A90563.1 GI:6739077
 KEYWORDS
 SOURCE unidentified.
 ORGANISM unidentified.
 REFERENCE 1 (bases 1 to 14)
 AUTHORS Brysch,W.D. and Schlingsenslepen,K.D.
 TITLE An antisense oligonucleotide preparation method
 JOURNAL Patent: EP 0856579-A 05-AUG-1998;
 BIOGNOSTIK GES (DE)
 FEATURES Location/Qualifiers
 source 1..14
 /organism="unidentified"
 /db_xref="taxon:32644"
 BASE COUNT 5 a 1 c 1 g 7 t
 ORIGIN

Query Match 100.0%; Score 7; DB 81; Length 14;
 Best Local Similarity 100.0%; Pred. No. 6.2e+05;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtata 7
 |||||
 Db 1 TTTGATA 7

RESULT 9
 AR082368 AR082368 14 bp DNA PAT 31-AUG-2000
 LOCUS Sequence 212 from patent US 5972704.
 DEFINITION AR082368
 ACCESSION AR082368
 VERSION AR082368.1 GI:10009094
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unclassified.
 REFERENCE 1 (bases 1 to 14)
 AUTHORS Draper,K.G., Chowrira,B., McSwiggen,J., Stinchcomb,D.T. and

Thompson, J.D.
HIV nef targeted ribozymes
JOURNAL Patent: US 5972704-A 212 26-OCT-1999;
FEATURES Location/Qualifiers
source 1. .14
BASE COUNT 5 a 1 c 4 g 4 t
ORIGIN

Query Match 100.0%; Score 7; DB 81; Length 14;
Best Local Similarity 100.0%; Pred. No. 6.2e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ttgtgata 7
|||||
Db 3 TTTGATA 9

RESULT 10
LOCUS I78414 14 bp DNA PAT 03-APR-1998
DEFINITION Sequence 212 from patent US 5693535.
ACCESSION I78414
VERSION I78414.1 GI:3014568
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 14)
AUTHORS Draper, K.G., Chowrira, B., McSwiggen, J., Stinchcomb, D.T. and Thompson, J.D.
TITLE HIV targeted ribozymes
JOURNAL Patent: US 5693535-A 212 02-DEC-1997;
FEATURES Location/Qualifiers
source 1. .14
BASE COUNT 5 a 1 c 4 g 4 t
ORIGIN

Query Match 100.0%; Score 7; DB 82; Length 14;
Best Local Similarity 100.0%; Pred. No. 6.2e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ttgtgata 7
|||||
Db 3 TTTGATA 9

RESULT 11
LOCUS AX007638/c 15 bp DNA VRL 06-SEP-2000
DEFINITION Sequence 180 from Patent WO9967428.
ACCESSION AX007638
VERSION AX007638.1 GI:9995335
KEYWORDS
SOURCE Aids-associated retrovirus.
ORGANISM Aids-associated retrovirus
REFERENCE 1 (bases 1 to 15)
AUTHORS Stuyver, L.
TITLE Method for detection of drug-selected mutations in the hiv protease gene
JOURNAL Patent: WO 9967428-A 29-DEC-1999;
FEATURES Location/Qualifiers
source 1. .15
BASE COUNT 5 a 1 c 3 g 6 t
ORIGIN

Query Match 100.0%; Score 7; DB 81; Length 15;
Best Local Similarity 100.0%; Pred. No. 6.1e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ttgtgata 7
|||||
Db 12 TTTGATA 6

RESULT 12
LOCUS AX007675/c 15 bp DNA VRL 06-SEP-2000
DEFINITION Sequence 217 from Patent WO9967428.
ACCESSION AX007675
VERSION AX007675.1 GI:9995372
KEYWORDS
SOURCE Aids-associated retrovirus.
ORGANISM Aids-associated retrovirus
REFERENCE 1 (bases 1 to 15)
AUTHORS Stuyver, L.
TITLE Method for detection of drug-selected mutations in the hiv protease gene
JOURNAL Patent: WO 9967428-A 29-DEC-1999;
FEATURES Location/Qualifiers
source 1. .15
BASE COUNT 5 a 1 c 4 g 5 t
ORIGIN

Query Match 100.0%; Score 7; DB 79; Length 15;
Best Local Similarity 100.0%; Pred. No. 6.1e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ttgtgata 7
|||||
Db 12 TTTGATA 6

RESULT 13
LOCUS A35660/c 15 bp DNA PAT 02-DEC-1996
DEFINITION Synthetic human IFN-alpha 2 gene oligo.
ACCESSION A35660
VERSION A35660.1 GI:1927042
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct
REFERENCE 1 (bases 1 to 15)
AUTHORS Camble, R. and Edge, M.D.
TITLE Analogous interferon polypeptides, process for their preparation and pharmaceutical compositions containing them
JOURNAL Patent: EP 0194006-A 105 10-SEP-1986;
FEATURES Location/Qualifiers
source 1. .15
BASE COUNT 5 a 4 c 2 g 4 t
ORIGIN

Query Match 100.0%; Score 7; DB 81; Length 15;
Best Local Similarity 100.0%; Pred. No. 6.1e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ttgtgata 7
|||||

Db 12 TTGTGATA 6

RESULT 14
AR045297/C
LOCUS AR045297 15 bp DNA PAT 29-SEP-1999
DEFINITION Sequence 90 from patent US 5817796.
ACCESSION AR045297
VERSION AR045297.1 GI:5966762
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 15)
Unclassified.
AUTHORS Stinchcomb,D.T., Draper,K., McSwiggen,J. and Jarvis,T.
TITLE C-myb ribozymes having 2'-5'-linked adenylyate residues
JOURNAL Patent: US 5817796-A 90 06-OCT-1998;
FEATURES
Location/Qualifiers
1..15
source
/organism="unknown"
BASE COUNT 6 a 3 c 3 g 3 t
ORIGIN

Query Match 100.0%; Score 7; DB 81; Length 15;
Best Local Similarity 100.0%; Pred. No. 6.1e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtata 7
|||||||
Db 11 TTGTGATA 5

RESULT 15
AR071525/C
LOCUS AR071525 15 bp DNA PAT 18-FEB-2000
DEFINITION Sequence 25 from patent US 5911982.
ACCESSION AR071525
VERSION AR071525.1 GI:7222413
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 15)
Unclassified.
AUTHORS Chao,Y.
TITLE Hz-1 virus persistence-associated-gene 1 (PAG1) promoter uses
therefor, and compositions containing same or products therefrom
JOURNAL Patent: US 5911982-A 25 15-JUN-1999;
FEATURES
Location/Qualifiers
1..15
source
/organism="unknown"
BASE COUNT 9 a 2 c 2 g 2 t
ORIGIN

Query Match 100.0%; Score 7; DB 81; Length 15;
Best Local Similarity 100.0%; Pred. No. 6.1e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtata 7
|||||||
Db 12 TTGTGATA 6

Search completed: March 27, 2001, 08:17:04
Job time: 5901 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 27, 2001, 07:38:20 ; Search time 2517.78 Seconds
(without alignments)
19.482 Million cell updates/sec

Title: US-09-380-826A-4
Perfect score: 7
Sequence: 1 tgttga 7

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues

Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST: *
1: gb_estl: *
2: gb_est2: *
3: gb_est3: *
4: gb_est4: *
5: gb_est5: *
6: gb_est6: *
7: gb_est7: *
8: gb_est8: *
9: gb_est9: *
10: gb_est10: *
11: gb_est11: *
12: gb_est12: *
13: gb_est13: *
14: gb_est14: *
15: gb_est15: *
16: gb_est16: *
17: gb_est17: *
18: gb_est18: *
19: gb_est19: *
20: gb_est20: *
21: gb_est21: *
22: gb_est22: *
23: gb_est23: *
24: gb_est24: *
25: gb_est25: *
26: gb_est26: *
27: gb_est27: *
28: gb_est28: *
29: gb_est29: *
30: gb_est30: *
31: gb_est31: *
32: gb_est32: *
33: gb_est33: *
34: gb_est34: *
35: gb_est35: *
36: gb_est36: *
37: gb_est37: *
38: gb_est38: *
39: gb_est39: *
40: gb_est40: *
41: em_estba: *
42: em_estfun: *
43: em_esthuml: *

44: em_esthum2: *
45: em_esthum3: *
46: em_esthum4: *
47: em_esthum5: *
48: em_esthum6: *
49: em_esthum7: *
50: em_esthum8: *
51: em_esthum9: *
52: em_esthum10: *
53: em_esthum11: *
54: em_esthum12: *
55: em_esthum13: *
56: em_esthum14: *
57: em_esthum15: *
58: em_esthum16: *
59: em_esthum17: *
60: em_esthum18: *
61: em_esthum19: *
62: em_esthum20: *
63: em_estin1: *
64: em_estin2: *
65: em_estin3: *
66: em_estin4: *
67: em_estov1: *
68: em_estov2: *
69: em_estpl1: *
70: em_estpl2: *
71: em_estpl3: *
72: em_estpl4: *
73: em_estpl5: *
74: em_estrol: *
75: em_estro2: *
76: em_estro3: *
77: em_estro4: *
78: em_estro5: *
79: em_estro6: *
80: em_estro7: *
81: em_estro8: *
82: em_estro9: *
83: em_estro10: *
84: em_estro11: *
85: em_estro12: *
86: em_estro13: *
87: gb_est41: *
88: gb_est42: *
89: gb_est43: *
90: gb_est44: *
91: gb_est45: *
92: gb_est46: *
93: gb_est47: *
94: gb_est48: *
95: gb_est49: *
96: gb_est50: *
97: gb_est51: *
98: gb_est52: *
99: gb_est53: *
100: gb_est54: *
101: gb_est55: *
102: gb_est56: *
103: gb_est57: *
104: gb_est67: *
105: gb_est68: *
106: gb_est69: *
107: gb_est70: *
108: gb_est71: *
109: gb_est72: *
110: gb_est73: *
111: gb_est74: *
112: em_esthum21: *
113: em_esthum22: *
114: em_esthum23: *
115: em_estom1: *
116: em_estom2: *

190: gb_gss25:.*
191: gb_gss26:.*
192: gb_gss27:.*
193: gb_gss28:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	7	100.0	19	171	AZ386406	IM0145C22
2	7	100.0	20	173	AZ484701	IM0311C24
3	7	100.0	21	171	AZ361153	IM0104A17
4	7	100.0	21	172	AZ450829	IM0249E13
5	7	100.0	21	174	AZ508369	IM0350A24
6	7	100.0	25	15	AI074857	oy27b11.s
7	7	100.0	25	16	AI158948	AI158948 u142807.x
8	7	100.0	25	20	AI453394	AI453394 t37802.x
9	7	100.0	26	171	AZ352012	IM0090M13
10	7	100.0	29	170	AZ332711	IM0061N14
11	7	100.0	29	172	AZ419519	IM0196N03
12	7	100.0	31	170	AZ333172	IM0062B09
13	7	100.0	31	171	AZ357647	IM0099D19
14	7	100.0	32	170	AZ320254	IM0040P07
15	7	100.0	32	170	AZ328463	IM0052D18
16	7	100.0	32	173	AZ467843	IM0279F06
17	7	100.0	33	141	H48291	H48291 yq77a01.r1
18	7	100.0	34	172	AZ408169	IM0179C17
19	7	100.0	35	170	AZ320436	IM0040C03
20	7	100.0	35	170	AZ322327	IM0043P14
21	7	100.0	36	148	AQ026010	EP(2)0461
22	7	100.0	36	170	AZ328537	IM0052B23
23	7	100.0	36	172	AZ433353	IM0219J12
24	7	100.0	37	1	AA011782	AA011782 mg92b03.r
25	7	100.0	37	10	AA636499	AA636499 vr18f02.r
26	7	100.0	37	140	D20685	D20685 HUMGS01661
27	7	100.0	37	145	T70764	T70764 yd14602.r1
28	7	100.0	37	148	AQ025921	1(2)k1010
29	7	100.0	38	113	HSM007685	AL042835 Homo sapi
30	7	100.0	38	173	AZ495773	AZ495773 IM0331O13
31	7	100.0	39	14	AI001077	AI001077 os66d04.s
32	7	100.0	39	173	AZ493504	AZ493504 IM0328P12
33	7	100.0	40	13	AA871171	AA871171 vq32a05.r
34	7	100.0	40	22	AI588805	AI588805 fb99f01.y
35	7	100.0	40	28	AU009892	AU009892 AU009892
36	7	100.0	40	28	AU009893	AU009893 AU009893
37	7	100.0	40	28	AU009894	AU009894 AU009894
38	7	100.0	40	28	AU009895	AU009895 AU009895
39	7	100.0	40	28	AU010246	AU010246 AU010246
40	7	100.0	40	28	AU010247	AU010247 AU010247
41	7	100.0	40	172	AZ447227	AZ447227 IM0244J13
42	7	100.0	40	173	AZ458268	AZ458268 IM0262N11
43	7	100.0	41	140	D25852	D25852 HUMGS04228
44	7	100.0	41	170	AZ324505	AZ324505 IM0046B02
45	7	100.0	41	171	AZ361377	AZ361377 IM0106J05

ALIGNMENTS

RESULT	1					
AZ386406	AZ386406	19 bp	DNA	GSS	02-OCT-2000	
LOCUS	IM0145C22F	Mouse 10kb plasmid	UUGC1M	library	Mus musculus	genomic
DEFINITION	clone UUGC1M0145C22 F	DNA sequence.				
ACCESSION	AZ386406					
VERSION	AZ386406.1	GI:10500106				
KEYWORDS	GSS.					
SOURCE	house mouse.					


```

ORGANISM      Mus musculus
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS        Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
                1 (bases 1 to 19)
                Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
                Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
                ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
                and Wright,D., Weiss,R.
TITLE          Mouse whole genome scaffolding with paired end reads from 10kb
                plasmid inserts
JOURNAL        Unpublished (2000)
COMMENT        Contact: Robert B. Weiss
                University of Utah Genome Center
                University of Utah
                Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
                84112, USA
                Tel: 801 585 5606
                Fax: 801 585 7177
                Email: ddunn@genetics.utah.edu
                Insert Length: 10000 Std Error: 0.00
                Plate: 0145 row: C column: 22
                Seq primer: CGTTGTAAACGACGCCAGT
                Class: plasmid ends
                High quality sequence stop: 19.
FEATURES       Location/Qualifiers
                1..19
                /organism="Mus musculus"
                /strain="C57BL/6J"
                /db_xref="taxon:10090"
                /clone="UUGC1M0145C22"
                /clone_lib="Mouse 10kb plasmid UUGC1M library"
                /sex="Male"
                /lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
                /note="Vector: PWD42nv; Purified genomic DNA from M.
                musculus C57BL/6J (male) was obtained from the Jackson
                Laboratory Mouse DNA Resource
                (http://www.jax.org/resources/documents/dnares/). The DNA
                was hydrodynamically sheared by repeated passage through a
                0.005 inch orifice at constant velocity. The sheared DNA
                was blunt end-repaired with T4 DNA polymerase and T4
                polynucleotide kinase. Adaptor oligonucleotides were
                ligated to the blunt ends in high molar excess. The
                adapted DNA was purified and size-selected for a 9.5 to
                10.5 kb range using preparative agarose gel
                electrophoresis. Vector DNA was prepared from a derivative
                of pWD42 (gil4732114[gb|AF129072.1]), a copy-number
                inducible derivative of plasmid R1. The vector was ligated
                with adaptors complementary to the insert adaptors and
                purified. The sheared, adapted mouse DNA was annealed to
                adapted vector DNA, and transformed into
                chemically-competent E. coli XL10-Gold (Stratagene) cells
                and selected for ampicillin resistance."
                5 a 2 c 6 g 6 t
BASE COUNT
ORIGIN
                Query Match 100.0%; Score 7; DB 171; Length 19;
                Best Local Similarity 100.0%; Pred. No. 1.5e+05;
                Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
                Qy 1 tgttggg 7
                Db 11 TGTGGG 17
RESULT 2
LOCUS      AZ484701 20 bp DNA GSS 05-OCT-2000
DEFINITION IM0311C24F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC1M0311C24 F, DNA sequence.
ACCESSION AZ484701
VERSION    AZ484701.1 GI:10649799
KEYWORDS   GSS.

```

```

SOURCE      house mouse.
ORGANISM    Mus musculus
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
                1 (bases 1 to 20)
                Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
                Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
                ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
                and Wright,D., Weiss,R.
TITLE        Mouse whole genome scaffolding with paired end reads from 10kb
                plasmid inserts
JOURNAL      Unpublished (2000)
COMMENT      Contact: Robert B. Weiss
                University of Utah Genome Center
                University of Utah
                Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
                84112, USA
                Tel: 801 585 5606
                Fax: 801 585 7177
                Email: ddunn@genetics.utah.edu
                Insert Length: 10000 Std Error: 0.00
                Plate: 0311 row: C column: 24
                Seq primer: CGTTGTAAACGACGCCAGT
                Class: plasmid ends
                High quality sequence stop: 20.
FEATURES     Location/Qualifiers
                1..20
                /organism="Mus musculus"
                /strain="C57BL/6J"
                /db_xref="taxon:10090"
                /clone="UUGC1M0311C24"
                /clone_lib="Mouse 10kb plasmid UUGC1M library"
                /sex="Male"
                /lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
                /note="Vector: PWD42nv; Purified genomic DNA from M.
                musculus C57BL/6J (male) was obtained from the Jackson
                Laboratory Mouse DNA Resource
                (http://www.jax.org/resources/documents/dnares/). The DNA
                was hydrodynamically sheared by repeated passage through a
                0.005 inch orifice at constant velocity. The sheared DNA
                was blunt end-repaired with T4 DNA polymerase and T4
                polynucleotide kinase. Adaptor oligonucleotides were
                ligated to the blunt ends in high molar excess. The
                adapted DNA was purified and size-selected for a 9.5 to
                10.5 kb range using preparative agarose gel
                electrophoresis. Vector DNA was prepared from a derivative
                of pWD42 (gil4732114[gb|AF129072.1]), a copy-number
                inducible derivative of plasmid R1. The vector was ligated
                with adaptors complementary to the insert adaptors and
                purified. The sheared, adapted mouse DNA was annealed to
                adapted vector DNA, and transformed into
                chemically-competent E. coli XL10-Gold (Stratagene) cells
                and selected for ampicillin resistance."
                2 a 4 c 6 g 8 t
BASE COUNT
ORIGIN
                Query Match 100.0%; Score 7; DB 173; Length 20;
                Best Local Similarity 100.0%; Pred. No. 1.5e+05;
                Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
                Qy 1 tgttggg 7
                Db 8 TGTGGG 14
RESULT 3
LOCUS      AZ361153 21 bp DNA GSS 02-OCT-2000
DEFINITION IM0104A17R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC1M0104A17 R, DNA sequence.
ACCESSION AZ361153
VERSION    AZ361153.1 GI:10474853

```

KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 21)
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C.,
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly
 M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.
 and Wright, D., Weiss, R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0104 row: A column: 17
 Seq primer: CACACAGAACACCTATGACC
 Class: plasmid ends
 High quality sequence stop: 21.
FEATURES
 Location/Qualifiers
 1..21
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUCG1M010417"
 /clone_lib="Mouse 10kb plasmid UUCG1M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
 /note="Vector: PWD42nv; Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of pWD42 (g114732114|gb|AF129072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."
BASE COUNT 10 a 5 c 0 g 6 t
ORIGIN
 Query Match 100.0%; Score 7; DB 171; Length 21;
 Best Local Similarity 100.0%; Pred. No. 1.5e+05;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 tgttggg 7
 Db 13 TGTGGA 7
RESULT 4
 AZ450829/c 21 bp DNA GSS 04-OCT-2000
 LOCUS
 DEFINITION 1M0249E13R Mouse 10kb plasmid UUCG1M library Mus musculus genomic
 clone UUCG1M0249E13 R, DNA sequence.
 AZ450829
 AZ450829

AZ450829.1 GI:10606020
GSS.
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 21)
AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C.,
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly
 M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.
 and Wright, D., Weiss, R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0249 row: E column: 13
 Seq primer: CACACAGAACACCTATGACC
 Class: plasmid ends
 High quality sequence stop: 21.
FEATURES
 Location/Qualifiers
 1..21
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUCG1M0249E13"
 /clone_lib="Mouse 10kb plasmid UUCG1M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
 /note="Vector: PWD42nv; Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of pWD42 (g114732114|gb|AF129072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."
BASE COUNT 8 a 8 c 0 g 5 t
ORIGIN
 Query Match 100.0%; Score 7; DB 172; Length 21;
 Best Local Similarity 100.0%; Pred. No. 1.5e+05;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 tgttggg 7
 Db 19 TGTGGA 13
RESULT 5
 AZ508369 21 bp DNA GSS 05-OCT-2000
 LOCUS
 DEFINITION 1M0350A24R Mouse 10kb plasmid UUCG1M library Mus musculus genomic
 clone UUCG1M0350A24 R, DNA sequence.
 AZ508369

[illegible]

```

AZ508369          AZ508369.1   GI:10689781
ACCESSION         GSS.
VERSION           house mouse.
KEYWORDS          Mus musculus
SOURCE            Mammalia; Eutheraia; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
ORGANISM          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  1 (bases 1 to 21)
REFERENCE         Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
AUTHORS           Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
                  ,M., Rose,R., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
TITLE             and Wright,D., Weiss R.
                  Mouse whole genome scaffolding with paired end reads from 10kb
JOURNAL           plasmid inserts
COMMENT           Unpublished (2000)
                  Contact: Robert B. Weiss
                  University of Utah Genome Center
                  Em.: 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
                  84112, USA
                  Tel: 801 585 5606
                  Fax: 801 585 7177
                  Email: dunn@genetics.utah.edu
                  Insert Length: 10000 Std Error: 0.00
                  Plate: 0350 row: A column: 24
                  Seq primer: CACAGCAAGAAACAGCTATGACC
                  Class: plasmid ends
                  High quality sequence stop: 21.
FEATURES          Location/Qualifiers
source            1..21
                  /organism="Mus musculus"
                  /strain="C57BL/6J"
                  /db_xref="taxon:10090"
                  /clone="UUCGLW0350A24"
                  /clone_lib="Mouse 10kb plasmid UUCCLM library"
                  /sex="Male"
                  /lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
                  /note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PWD42 (gi|4732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid RI. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT        5 a      7 g      8 t
ORIGIN            1 c
Query Match       100.0%; Score 7; DB 174; Length 21;
Best Local Similarity 100.0%; Pred. No. 1.5e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy               1 ttgttga 7
                |||||||
Db               3 TGTTTGA 9
RESULT           6
LOCUS            AI074857       25 bp     mRNA                      EST
DEFINITION       oy27b11.s1 Soares_senescent_fibroblasts_Nbhsf Homo sapiens CDNA
REFERENCE
AUTHORS          Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.
                  Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
                  Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
                  24-SEP-1998

```

Wed Mar 28 14:03:19 2001

Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.

The WashU-HHMI Mouse EST Project

Unpublished (1996)

Contact: Maria M/Mouse EST Project

WashU-HHMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

MG1:969416

Trace considered overall poor quality

Possible reversed clone: similarity on wrong strand

Seq primer: custom primer used

High quality sequence stop: 1.

FEATURES
source

1. .25
/organism="Mus musculus"
/strain="C57BL"
/db_xref="taxon:10090"
/clone="IMAGE:1885092"
/clone_lib="Sugano mouse embryo mewa"
/dev_stage="embryo, 14 dpc"
/lab_host="DH10B"

/note="Vector: pME18S-FL3; Site_1: DraIII (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed with an oligo(dT) primer (ATGCGCCCTTTTTTTTTTTTTT); double-stranded cDNA was ligated to a DraIII adaptor [TGTGGCCCTACTG], digested and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CTTCTGCTCTAAAGCTGCG and 3' end primer CGACCTGACGCTGAGGACA."

BASE COUNT 7 a 5 c 5 g 8 t

ORIGIN

Query Match 100.0%; Score 7; DB 16; Length 25;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgttga 7

Db 9 TGTGGA 15

RESULT 8
AI453394/c

LOCUS

DEFINITION AI453394 25 bp mRNA EST 13-APR-1999
t37902.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2143730 3' similar to TR:Q39949 Q39949 HYDROXYPROLINE-RICH PROTEIN.; contains element MSK1 repetitive element.; mRNA sequence.

ACCESSION AI453394

VERSION AI453394.1

KEYWORDS GI:4281559

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

TITLE 1 (bases 1 to 25)

JOURNAL NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Life Technologies catalog #: 11548-013

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Trace considered overall poor quality

Insert Length: 1074 Std Error: 0.00

Seq primer: 400P from Gibco

High quality sequence stop: 1.

FEATURES
source

1. .25

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:2143730"

/clone_lib="NCI_CGAP_Pan1"

/tissue_type="adenocarcinoma"

/lab_host="DH10B"

/note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: SalI;

Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.

Average insert size 1.72 kb. Life Technologies catalog #:

11548-013"

BASE COUNT 8 a 16 c 0 g 1 t

ORIGIN

Query Match 100.0%; Score 7; DB 20; Length 25;

Best Local Similarity 100.0%; Pred. No. 1.6e+05;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgttga 7

Db 8 TGTGGA 2

RESULT 9

AZ352012/c

LOCUS

DEFINITION AZ352012 26 bp DNA GSS 29-SEP-2000

clone UUGC1M0090M13 F, DNA sequence.

ACCESSION AZ352012

VERSION AZ352012.1

KEYWORDS GI:10431249

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 26)

Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,

Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly

,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.

and Wright,D., Weiss,R.

Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah Genome Center

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0090 row: M column: 13

Seq primer: CTTGTAAACGACGCCAGT

Class: plasmid ends

High quality sequence stop: 26.

Location/Qualifiers

1. 26

/organism="Mus musculus"

/strain="C57BL/6J"

```

/db_xref="taxon:10090"
/clone="UUGC1M0090M13"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/notes="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT      8 a      10 c      2 g      6 t
ORIGIN

Query Match      100.0%; Score 7; DB 171; Length 26;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgttggg 7
    |||||
Db 19 TGTGGGA 13

RESULT 10
A2332711
LOCUS      29 bp      DNA      GSS      29-SEP-2000
DEFINITION clone UUGC1M0061N14 F, DNA sequence.
ACCESSION  A2332711
VERSION     A2332711.1 GI:10396624
KEYWORDS   GSS.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 29)
AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
JOURNAL   Unpublished (2000)
COMMENT   Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000 Std Error: 0.00
            Plate: 0061 row: N column: 14
            Seq primer: CTTGTAAACGACGCCAGT
            Class: plasmid ends
            High quality sequence stop: 29.
            Location/Qualifiers
                1..29
                /organism="Mus musculus"
FEATURES
source
1..29
location/Qualifiers
1..29

```

```

/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0061N14"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/notes="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT      5 a      4 c      9 g      11 t
ORIGIN

Query Match      100.0%; Score 7; DB 170; Length 29;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgttggg 7
    |||||
Db 3 TGTGGGA 9

RESULT 11
A2419519/c
LOCUS      29 bp      DNA      GSS      03-OCT-2000
DEFINITION clone UUGC1M0196N03 F, DNA sequence.
ACCESSION  A2419519
VERSION     A2419519.1 GI:10543532
KEYWORDS   GSS.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 29)
AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
JOURNAL   Unpublished (2000)
COMMENT   Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000 Std Error: 0.00
            Plate: 0196 row: N column: 03
            Seq primer: CTTGTAAACGACGCCAGT
            Class: plasmid ends
            High quality sequence stop: 29.
            Location/Qualifiers
                1..29
                /organism="Mus musculus"
FEATURES
source
1..29
location/Qualifiers
1..29

```

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0196N03"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/notes="Vector: pWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gil4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT      9 a      5 g      7 t
ORIGIN
11 a      12 c      4 g      4 t

Query Match      100.0%; Score 7; DB 172; Length 29;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tggtaga 7
    |||||
Db 17 TGTGGA 11

RESULT 12
AZ333172/c
LOCUS      31 bp      DNA      29-SEP-2000
DEFINITION LM0062B09F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC1M0062B09 F, DNA sequence.
ACCESSION  AZ333172
VERSION     AZ333172.1  GI:10397530
KEYWORDS   GSS.
SOURCE     house mouse.
ORGANISM   Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0062 row: B column: 09
Seq primer: CGTTGTAACGACGCCAGT
Class: plasmid ends
High quality sequence stop: 31.
Location/Qualifiers

source
1. .31
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0062B09"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/notes="Vector: pWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gil4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT      11 a      12 c      4 g      4 t
ORIGIN

Query Match      100.0%; Score 7; DB 170; Length 31;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tggtaga 7
    |||||
Db 14 TGTGGA 8

RESULT 13
AZ357647/c
LOCUS      31 bp      DNA      02-OCT-2000
DEFINITION LM0099D19F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC1M0099D19 F, DNA sequence.
ACCESSION  AZ357647
VERSION     AZ357647.1  GI:10471347
KEYWORDS   GSS.
SOURCE     house mouse.
ORGANISM   Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0099 row: D column: 19
Seq primer: CGTTGTAACGACGCCAGT
Class: plasmid ends
High quality sequence stop: 31.

```

FEATURES
source

Location/Qualifiers
1. .31
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGCLM0099D19"
/clone_lib="Mouse 10kb plasmid UUGCLM library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 10 a 12 c 5 g 4 t
ORIGIN

Query Match 100.0%; Score 7; DB 171; Length 31;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgttga 7
Db 13 TGTGGA 7

RESULT 14

AZ320254/c
LOCUS AZ320254 32 bp DNA GSS 29-SEP-2000
DEFINITION clone UUGCLM0040P07 F, DNA sequence.

ACCESSION AZ320254
VERSION 1
KEYWORDS GSS.

SOURCE house mouse.

ORGANISM

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 32)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

JOURNAL

COMMENT Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0040 row: P column: 07

Seq primer: CGTTGTAAACGCGCCAGT

Class: plasmid ends

FEATURES
source

High quality sequence stop: 32.
Location/Qualifiers
1. .32

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGCLM0040P07"
/clone_lib="Mouse 10kb plasmid UUGCLM library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 9 a 11 c 3 g 9 t
ORIGIN

Query Match 100.0%; Score 7; DB 170; Length 32;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgttga 7
Db 22 TGTGGA 16

RESULT 15

AZ328463

LOCUS

DEFINITION AZ328463 32 bp DNA GSS 29-SEP-2000
clone UUGCLM0052D18 F, DNA sequence.

ACCESSION AZ328463
VERSION 1
KEYWORDS GSS.

SOURCE house mouse.

ORGANISM

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 32)

REFERENCE

AUTHORS

Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

JOURNAL

COMMENT

Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0052 row: D column: 18

Seq primer: CGTTGTAAACGCGCCAGT

Class: plasmid ends
High quality sequence stop: 32.

FEATURES
Location/Qualifiers
1..32
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0052D18"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (g11473211419b1AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 7 a 3 c 12 g 10 t
ORIGIN

Query Match 100.0%; Score 7; DB 170; Length 32;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tqtgtgga 7
|||||
Db 7 TGTGTGA 13

Search completed: March 27, 2001, 07:38:21
Job time: 4584 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 27, 2001, 07:38:17 ; Search time 2517.78 Seconds
(without alignments)
61.230 Million cell updates/sec

Title: US-09-380-826A-2
Perfect score: 22
Sequence: 1 ttgtggatcacaaagtattgata 22

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues

Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:

- 1: gb_est1.*
- 2: gb_est2.*
- 3: gb_est3.*
- 4: gb_est4.*
- 5: gb_est5.*
- 6: gb_est6.*
- 7: gb_est7.*
- 8: gb_est8.*
- 9: gb_est9.*
- 10: gb_est10.*
- 11: gb_est11.*
- 12: gb_est12.*
- 13: gb_est13.*
- 14: gb_est14.*
- 15: gb_est15.*
- 16: gb_est16.*
- 17: gb_est17.*
- 18: gb_est18.*
- 19: gb_est19.*
- 20: gb_est20.*
- 21: gb_est21.*
- 22: gb_est22.*
- 23: gb_est23.*
- 24: gb_est24.*
- 25: gb_est25.*
- 26: gb_est26.*
- 27: gb_est27.*
- 28: gb_est28.*
- 29: gb_est29.*
- 30: gb_est30.*
- 31: gb_est31.*
- 32: gb_est32.*
- 33: gb_est33.*
- 34: gb_est34.*
- 35: gb_est35.*
- 36: gb_est36.*
- 37: gb_est37.*
- 38: gb_est38.*
- 39: gb_est39.*
- 40: gb_est40.*
- 41: em_estba.*
- 42: em_estfun.*
- 43: em_esthum1.*

- 44: em_esthum2.*
- 45: em_esthum3.*
- 46: em_esthum4.*
- 47: em_esthum5.*
- 48: em_esthum6.*
- 49: em_esthum7.*
- 50: em_esthum8.*
- 51: em_esthum9.*
- 52: em_esthum10.*
- 53: em_esthum11.*
- 54: em_esthum12.*
- 55: em_esthum13.*
- 56: em_esthum14.*
- 57: em_esthum15.*
- 58: em_esthum16.*
- 59: em_esthum17.*
- 60: em_esthum18.*
- 61: em_esthum19.*
- 62: em_esthum20.*
- 63: em_estin1.*
- 64: em_estin2.*
- 65: em_estin3.*
- 66: em_estin4.*
- 67: em_estov1.*
- 68: em_estov2.*
- 69: em_estpl1.*
- 70: em_estpl2.*
- 71: em_estpl3.*
- 72: em_estpl4.*
- 73: em_estpl5.*
- 74: em_estro1.*
- 75: em_estro2.*
- 76: em_estro3.*
- 77: em_estro4.*
- 78: em_estro5.*
- 79: em_estro6.*
- 80: em_estro7.*
- 81: em_estro8.*
- 82: em_estro9.*
- 83: em_estro10.*
- 84: em_estro11.*
- 85: em_estro12.*
- 86: em_estro13.*
- 87: gb_est41.*
- 88: gb_est42.*
- 89: gb_est43.*
- 90: gb_est44.*
- 91: gb_est45.*
- 92: gb_est46.*
- 93: gb_est47.*
- 94: gb_est48.*
- 95: gb_est49.*
- 96: gb_est50.*
- 97: gb_est51.*
- 98: gb_est52.*
- 99: gb_est53.*
- 100: gb_est54.*
- 101: gb_est55.*
- 102: gb_est56.*
- 103: gb_est57.*
- 104: gb_est58.*
- 105: gb_est59.*
- 106: gb_est60.*
- 107: gb_est61.*
- 108: gb_est62.*
- 109: gb_est63.*
- 110: gb_est64.*
- 111: gb_est65.*
- 112: em_esthum21.*
- 113: em_esthum22.*
- 114: em_esthum23.*
- 115: em_estom1.*
- 116: em_estom2.*

117: em_estpl6:*
118: em_estpl7:*
119: em_estpl8:*
120: em_estpl9:*
121: em_estpl10:*
122: em_estpl11:*
123: em_estpl12:*
124: em_estpl13:*
125: em_estpl14:*
126: em_estpl15:*
127: em_estpl16:*
128: em_estpl17:*
129: em_estpl18:*
130: em_estpl19:*
131: em_estpl20:*
132: em_estpl21:*
133: em_estpl22:*
134: em_estpl23:*
135: em_estpl24:*
136: em_estpl25:*
137: em_estpl26:*
138: em_estpl27:*
139: em_estpl28:*
140: em_estpl29:*
141: em_estpl30:*
142: em_estpl31:*
143: em_estpl32:*
144: em_estpl33:*
145: em_estpl34:*
146: em_estpl35:*
147: em_estpl36:*
148: em_estpl37:*
149: em_estpl38:*
150: em_estpl39:*
151: em_estpl40:*
152: em_estpl41:*
153: em_estpl42:*
154: em_estpl43:*
155: em_estpl44:*
156: em_estpl45:*
157: em_estpl46:*
158: em_estpl47:*
159: em_estpl48:*
160: em_estpl49:*
161: em_estpl50:*
162: em_estpl51:*
163: em_estpl52:*
164: em_estpl53:*
165: em_estpl54:*
166: em_estpl55:*
167: em_estpl56:*
168: em_estpl57:*
169: em_estpl58:*
170: em_estpl59:*
171: em_estpl60:*
172: em_estpl61:*
173: em_estpl62:*
174: em_estpl63:*
175: em_estpl64:*
176: em_estpl65:*
177: em_estpl66:*
178: em_estpl67:*
179: em_estpl68:*
180: em_estpl69:*
181: em_estpl70:*
182: em_estpl71:*
183: em_estpl72:*
184: em_estpl73:*
185: em_estpl74:*
186: em_estpl75:*
187: em_estpl76:*
188: em_estpl77:*
189: em_estpl78:*

190: gb_gss25:*
191: gb_gss26:*
192: gb_gss27:*
193: gb_gss28:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	18.8	85.5	361	37	AV532040	AV532040 AV532040
C 2	18.8	85.5	394	147	Z26559	Z26559 ATTS1651 Ve
C 3	18.8	85.5	490	14	AB038725	AB038725 AB038725
C 4	18.8	85.5	527	37	AV520789	AV520789 AV520789
C 5	18.8	85.5	530	27	AI995637	AI995637 701676626
C 6	18.4	83.6	608	138	BE977618	BE977618 bs66808.y
C 7	18.4	83.6	934	191	CNS02BWA	AL190387 Tetradon
C 8	17.8	80.9	600	138	BF006444	BF006444 EST434942
C 9	17.8	80.9	633	138	BF006254	BF006254 EST434752
C 10	17.4	79.1	286	126	BB288727	BB288727 BB288727
C 11	17.4	79.1	407	110	BE581720	BE581720 kg51c07.y
C 12	17.4	79.1	529	158	AQ497779	AQ497779 HS_5066.B
C 13	17.4	79.1	674	94	AW695022	AW695022 NF082804S
C 14	17.4	79.1	762	106	BE283043	BE283043 601101332
C 15	17.4	79.1	837	110	BE642809	BE642809 Cr12_7.B0
C 16	17.4	79.1	1101	190	CNS00D26	AL077427 Drosoph11
C 17	17.2	78.2	151	134	BE068225	BE068225 MR4-BF036
C 18	17.2	78.2	425	15	A1049911	A1049911 an30h03.x
C 19	17.2	78.2	431	159	AQ595814	AQ595814 HS_2132.B
C 20	17.2	78.2	432	150	AQ221715	AQ221715 HS_2010.A
C 21	17.2	78.2	437	146	W43212	W43212 22591.LamBd
C 22	17.2	78.2	439	162	AQ796989	AQ796989 nbdx0071H
C 23	17.2	78.2	500	24	AI733664	AI733664 an30h03.x
C 24	17.2	78.2	524	174	B62582	B62582 T22F18TR.TA
C 25	17.2	78.2	531	150	AQ223477	AQ223477 HS_2003.B
C 26	17.2	78.2	536	39	AW034253	AW034253 EST377824
C 27	17.2	78.2	546	24	AI779714	AI779714 EST360593
C 28	17.2	78.2	579	174	B67515	B67515 T22M10TR.TA
C 29	17.2	78.2	602	24	AI777095	AI777095 EST358060
C 30	17.2	78.2	614	105	BE187570	BE187570 EST336131
C 31	17.2	78.2	619	97	AW963686	AW963686 EST375759
C 32	17.2	78.2	680	174	B57784	B57784 CIT-HSP-201
C 33	17.2	78.2	698	168	A2193715	A2193715 SP_1023.B
C 34	17.2	78.2	904	191	CNS029DN	AL187124 Tetradon
C 35	17.2	78.2	949	192	CNS04HNP	AL291166 Tetradon
C 36	17.2	77.3	373	140	C70416	C70416 C70416.Yuji
C 37	16.8	76.4	238	34	AV313332	AV313332 AV313332
C 38	16.8	76.4	260	104	BE118999	BE118999 UI-R-CA0-
C 39	16.8	76.4	296	145	T20362	T20362 6c01908-t7
C 40	16.8	76.4	396	97	AW943911	AW943911 LD47517.3
C 41	16.8	76.4	397	91	AW487226	AW487226 81727.MAR
C 42	16.8	76.4	432	157	AQ449764	AQ449764 500004C01
C 43	16.8	76.4	446	24	AI761541	AI761541 w161f07.x
C 44	16.8	76.4	450	138	BF002933	BF002933 7g51c09.x
C 45	16.8	76.4	464	156	AQ332265	AQ332265 HS_5007.A

ALIGNMENTS

RESULT 1
AV532040/c
LOCUS AV532040 Arabidopsis thaliana flower buds Columbia Arabidopsis.
DEFINITION thaliana cDNA clone FB034f09F 3', mRNA sequence.
ACCESSION AV532040
VERSION AV532040.1 GI:8692323
KEYWORDS EST
SOURCE thale cress.

```

ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 361)
AUTHORS Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
TITLE A large scale analysis of cDNA in Arabidopsis thaliana: Generation
of 12,028 non-redundant expressed sequence tags from normalized and
size-selected cDNA libraries
JOURNAL DNA Res. 7, 175-180 (2000)
MEDLINE 20363093
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES
    source
        1..361
            /organism="Arabidopsis thaliana"
            /strain="Columbia"
            /db_xref="taxon:3702"
            /clone="FB034f09p"
            /clone_lib="Arabidopsis thaliana flower buds Columbia"
            /tissue_type="flower buds"
            /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
            XhoI"
BASE COUNT 113 a 73 c 61 g 114 t
ORIGIN
Query Match 85.5%; Score 18.8; DB 37; Length 361;
Best Local Similarity 90.9%; Pred. No. 34;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 tgttgatcacaaagatttgata 22
||||| |||||||||
Db 276 TGTGGTTCACAGATTTCACA 255

RESULT 2
LOCUS 226559 394 bp mRNA EST 31-MAY-1995
DEFINITION ATTS1651 Versailles-VC Arabidopsis thaliana cDNA clone VCVDH08 3'
similar to Ribonuclease (RNS2), mRNA sequence.
ACCESSION Z26559
VERSION Z26559.1 GI:404227
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 394)
AUTHORS CNRS.
TITLE The Arabidopsis thaliana transcribed genome: the GDR cDNA program
JOURNAL Unpublished (1996)
COMMENT Contact: Desprez T., Anselem J., Chiapello H., Rouze P., Caboche
M., Hofte H.
INRA Versailles
Laboratoire de Biologie Cellulaire
Route de Saint-Cyr.78026 Versailles Cedex,France
Email: thierry@versailles.inra.fr.
Location/Qualifiers
    1..394
        /organism="Arabidopsis thaliana"
        /strain="ecotype Columbia"
        /db_xref="taxon:3702"
        /clone="VCVDH08"
        /clone_lib="Versailles-VC"
        /tissue_type="whole seedlings"
        /dev_stage="in vitro-grown etiolated seedlings,5 days old"
        /note="Vector: pBluescript"
BASE COUNT 117 a 80 c 63 g 133 t
1 others

```

```

ORIGIN
Query Match 85.5%; Score 18.8; DB 147; Length 394;
Best Local Similarity 90.9%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 tgttgatcacaaagatttgata 22
||||| |||||||||
Db 321 TGTGGTTCACAGATTTCACA 300

RESULT 3
LOCUS AB038725/c 490 bp mRNA EST 29-SEP-2000
DEFINITION AB038725 Arabidopsis thaliana Above-ground organ from two to
six-week old plants Columbia Arabidopsis thaliana cDNA clone
AP230q03_f 3', mRNA sequence.
ACCESSION AB038725
VERSION AB038725.1 GI:7212552
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 490)
AUTHORS Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
TITLE A large scale analysis of cDNA in Arabidopsis thaliana: Generation
of 12,028 non-redundant expressed sequence tags from normalized and
size-selected cDNA libraries
JOURNAL DNA Res. 7, 175-180 (2000)
MEDLINE 20363093
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
Location/Qualifiers
    1..490
        /organism="Arabidopsis thaliana"
        /strain="Columbia"
        /db_xref="taxon:3702"
        /clone="AP230q03_f"
        /clone_lib="Arabidopsis thaliana Above-ground organ from
        two to six-week old plants Columbia"
        /tissue_type="Above-ground organ from two to six-week old
        plants"
        /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
        XhoI"
BASE COUNT 147 a 99 c 87 g 157 t
ORIGIN
Query Match 85.5%; Score 18.8; DB 14; Length 490;
Best Local Similarity 90.9%; Pred. No. 36;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 tgttgatcacaaagatttgata 22
||||| |||||||||
Db 303 TGTGGTTCACAGATTTCACA 282

RESULT 4
LOCUS AV520789/c 527 bp mRNA EST 01-SEP-2000
DEFINITION AV520789 Arabidopsis thaliana aboveground organs two to six-week
old Arabidopsis thaliana cDNA clone AP232g06F 3', mRNA sequence.
ACCESSION AV520789
VERSION AV520789.1 GI:8680316
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana

```

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.

REFERENCE

AUTHORS

TITLE

1 (bases 1 to 527)
Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
A large scale analysis of cDNA in Arabidopsis thaliana: Generation of 12,028 non-redundant expressed sequence tags from normalized and size-selected cDNA libraries

JOURNAL

MEDLINE

COMMENT

20363093
Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizukazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES

source

1. 527
/location/Qualifiers
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone_lib="AP232g06f"
/clone_lib="Arabidopsi\$ thaliana aboveground organs two to six-week old"
/tissue_type="aboveground organs"
/dev_stage="two to six-week old"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 157 a 111 c 101 g 158 t
ORIGIN

Query Match 85.5%; Score 18.8; DB 37; Length 527;
Best Local Similarity 90.9%; Pred. No. 36;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 tgttgatcacagatttgata 22

||||| ||||||||| |||

Db 247 TGTGGTTCACAGATTGACA 226

RESULT 5
AI995637/c 530 bp mRNA EST 08-SEP-1999
LOCUS 701676626 A. thaliana, Columbia Col-0, inflorescence-1 Arabidopsi\$
DEFINITION thaliana cDNA clone 701676626, mRNA sequence.

ACCESSION AI995637

VERSION AI995637.1 GI:5842542

KEYWORDS EST.

SOURCE thale cress.

ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi\$.

REFERENCE

AUTHORS

TITLE

1 (bases 1 to 530)
Chen,J., Moniyama,M., Chan,E., Mooney,M., Carroon,B., Gilliland,D., Wang,X., Hillman,J., Guebler,K., Kim,C., Doyle,M., Brzoska,P., Gorgone,G., Burns,D., Griffin,J., Mouanoutoua,M., Nguyen,D., Tan,R., Rose,M., Warren,B., Ton,B., Kastury,K., Borillo,C., Carpio,T., Policky,J., Suzuki,G., Argentine,C., Shah,S., Nobriaga,A., Murry,L., Turner,C., Krikorian,S., Elder,L. and Hanson,D.

Arabidopsi\$ thaliana Gene Expression MicroArray

Unpublished (1999)

Contact: David Smoller, Ph.D.

Genome Systems, Inc., a wholly owned subsidiary of Incyte

Pharmaceuticals, Inc.

4633 World Parkway Circle, St. Louis, MO 63134, USA

Tel: 877-577-2733

Fax: 314-427-3324

Email: service@genomesystems.com.

location/Qualifiers

1. 530

/organism="Arabidopsis thaliana"

/cultivar="Columbia Col-0"

FEATURES

source

/db_xref="taxon:3702"

/clone_lib="A. thaliana, Columbia Col-0, inflorescence-1"

/tissue_type="inflorescence"

/dev_stage="4 - 7 weeks"

/note="Vector: pSPORT; Site_1: NotI; Site_2: SalI; cDNA library was derived from untreated inflorescence tissue from Arabidopsi\$ thaliana, Columbia Col-0, at 4 - 7 weeks. Plants were grown in 1:1:1 peat moss/vermiculite/perlite soil at 22 deg. C +/- 3 deg. C under constant light, and watered with fertilizer. cDNA synthesis was initiated using a NotI-oligo(dT) primer. Double-stranded cDNA was blunt-ended, ligated to SalI adaptors, digested with NotI, size-selected, and cloned into the NotI and SalI sites of the pSPORT vector."

BASE COUNT 163 a 108 c 100 g 158 t

ORIGIN

Query Match 85.5%; Score 18.8; DB 27; Length 530;
Best Local Similarity 90.9%; Pred. No. 36;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 tgttgatcacagatttgata 22

||||| ||||||||| |||

Db 254 TGTGGTTCACAGATTGACA 233

RESULT 6

BE977618/c 608 bp mRNA EST 04-OCT-2000

LOCUS bs66h08.y1 Drosophila melanogaster adult testis library Drosophilla

DEFINITION melanogaster cDNA clone bs66h08 5', mRNA sequence.

ACCESSION BE977618

VERSION BE977618.1 GI:10608272

KEYWORDS EST.

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophilla.

1 (bases 1 to 608)

Andrews, J., Bouffard, G. and Oliver, B.

Drosophila melanogaster testis expressed sequence tags

Unpublished (1999)

Contact: Brian Oliver

Laboratory of Cellular and Developmental Biology

NIDDK, National Institutes of Health

6 Center Drive MSC 2715, Bldg 6, Rm B1-13, Bethesda, MD 20892 USA

Fax: (301) 496 5239

Email: oliver@helix.nih.gov,

http://www.nidk.nih.gov/intram/people/boliver.htm

Tissue isolation and library construction performed at the National

Institute of Diabetes and Digestive and Kidney Diseases, NIH (see

http://www.nidk.nih.gov/intram/people/boliver.htm). DNA sequencing

and analyses performed by National Institutes of Health Intramural

Sequencing Center (NISC; see http://www.nisc.nih.gov).

Plate: 66 row: h column: 08

Seq primer: M13Kp1 reverse primer (ABI).

Location/Qualifiers

1. 608

/organism="Drosophila melanogaster"

/strain="y[*] w[67c1]/Y"

/db_xref="taxon:722"

/clone_lib="bs66h08"

/clone_lib="Drosophila melanogaster adult testis library"

/sex="male"

/dev_stage="1-5 day adult"

/lab_host="SOLR (Stratagene)"

/note="Organ: testis; Vector: pBluescript SK (Stratagene);

Site_1: EcoR I; Site_2: Xho I; Testes dissected from 1-5

day adult y[*] w[67c1]/Y males raised at 25oc. RNA

isolated using Trizol (Life Technologies) and a single

round of Poly(A)+ selection using Oligotex (Qiagen). cDNA library constructed using Stratagene ZAP-cDNA synthesis kit. Oligo dT-primed, size fractionated -1-6 kb, and directionally cloned at EcoRI and XhoI in Uni-ZAP XR. Following a single round of amplification pBluescript SK phagemids were mass excised. A distribution channel for clones is being sought, but not currently available. Requests for clones cannot be honored."

BASE COUNT 173 a 146 c 161 g 128 t
ORIGIN

Query Match 83.6%; Score 18.4; DB 138; Length 608;
Best Local Similarity 95.0%; Pred. NO. 59;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 ttggatcacaaagtattgata 22
||||| ||||||| |||||

Db 74 TTGGATCCCAAGATTGATA 55

RESULT 7

CNS02BWA/c

LOCUS 934 bp DNA GSS 12-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone 254C22 of library G from Tetraodon nigroviridis, genomic survey sequence.

ACCESSION AL190387

VERSION 1 GI:7828491

KEYWORDS GSS: genome survey sequence.

SOURCE Tetraodon nigroviridis.

ORGANISM Tetraodon nigroviridis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Eurypterygii; Ctenosquamata; Acanthomorpha; Euacanthomorpha; Holacanthopterygii; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodontidae; Tetraodon.

REFERENCE 1 (bases 1 to 934)

Roest-Crollius, H., Jallou, O., Dasilva, C., Fizes, C., Fisher, C., Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and Weissenbach, J.

Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis

Unpublished

2 (bases 1 to 934)

Roest-Crollius, H., Jallou, O., Dasilva, C., Bouneau, L., Fisher, C., Bernot, A., Fizes, C., Wincker, P., Brottier, P., Quetier, F., Saurin, W. and Weissenbach, J.

Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence

Unpublished

3 (bases 1 to 934)

Genoscope.

Direct Submission

Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases

This sequence is a single read and was generated as part of a large scale clone-and-sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/tetraodon>.

FEATURES

Location/Qualifiers

1..934

/organism="Tetraodon nigroviridis"

/db_xref="taxon:99883"

/clone="254C22"

/clone_lib="G"

/note="Genoscope sequence ID : COAG254BB11SP1-end ; PUC-Ori"

BASE COUNT 282 a 175 c 207 g 259 t 11 others

ORIGIN

Query Match 83.6%; Score 18.4; DB 191; Length 934;
Best Local Similarity 95.0%; Pred. NO. 62;

RESULT 9

BF006254/c

LOCUS 633 bp mRNA

DEFINITION EST434752 DLSLC Medicago truncatula cDNA clone pDLSLC-40G21, mRNA

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 3 ttggatcacaaagtattgata 22
||||| ||||||| |||||

Db 840 TTGGAACACAGATTGATA 821

RESULT 8

BF006444/c

LOCUS 600 bp mRNA

DEFINITION EST434942 DLSLC Medicago truncatula cDNA clone pDLSLC-41J15, mRNA

sequence.

ACCESSION BF006444

VERSION BF006444.1

KEYWORDS EST.

SOURCE barrel medic.

ORGANISM Medicago truncatula

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago.

REFERENCE 1 (bases 1 to 600)

Fedorova, M., Pierson, B.L., Samac, D.A., Ganitt, J.S., Vance, C.P., Gonzales, M.B. and Ellis, L.

ESTs from Medicago truncatula leaves and cotyledons

Unpublished (2000)

JOURNAL

COMMENT

Contact: Deborah A. Samac

Department of Plant Pathology

University of Minnesota

495 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108, USA

Tel: 612 625 1243

Fax: 651 649 5058

Email: debbyepuccini.crl.umn.edu

University of Minnesota name: M275482e TIGR sequence name:

MTLBA56TK More information is available at:

<http://chrysis.tamu.edu/medicago>

Seq primer: SKmod (CTA GAA CTA gta gAT CC).

Location/Qualifiers

1..600

/organism="Medicago truncatula"

/cultivar="genotype A17"

/db_xref="taxon:3880"

/clone="pDLSLC-41J15"

/clone_lib="DLSLC"

/tissue_type="leaves and cotyledons"

/dev_stage="mixture of cotyledons from five days old plants and leaves obtained from two weeks old plants"

/lab_host="E. coli strain SOLR"

/note="vector: pBluescript SK +/-; Site1: EcoRI; Site2: XhoI; cDNA was prepared from polyA+ enriched RNA from the mixture of cotyledons of five days old plants and leaves of two weeks old plants. The cDNA was directionally ligated into the Uni-ZAP XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-ZAP phage using Ex-Assist helper phage and propagated in SOLR cells."

BASE COUNT 127 a 120 c 132 g 221 t

ORIGIN

Query Match 80.9%; Score 17.8; DB 138; Length 600;
Best Local Similarity 90.5%; Pred. NO. 1.2e+02;

Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 tggatcacaaagtattgata 21

||||| ||||||| |||||

Db 90 TGTTCATCACAGATTAGAT 70

RESULT 9

BF006254/c

LOCUS 633 bp mRNA

DEFINITION EST434752 DLSLC Medicago truncatula cDNA clone pDLSLC-40G21, mRNA


```

RESULT 11
BE581720/c 407 bp mRNA EST 16-AUG-2000
LOCUS Kq51c07.y1 TBN95TM-SSR Strongyloides stercoralis cDNA 5' similar to
DEFINITION WP.T19810.2 CE16413 ;, mRNA sequence.
ACCESSION BE581720
VERSION BE581720.1 GI:9832662
KEYWORDS EST.
SOURCE Strongyloides stercoralis.
ORGANISM Strongyloides stercoralis.

REFERENCE
AUTHORS Eukaryota: Metazoa; Nematoda; Chromadorea; Rhabditida;
Panagrolaimoidae; Strongyloidea; Strongyloides.
1 (bases 1 to 407)
McCarter,J., Clifton,S., Marra,M., Hillier,L., Kucaba,T., Martin,J.,
Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen
M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey
N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas
M., McCann,R., Waterston,R. and Wilson,R. 1999
The Washington Univ. Nematode EST Project, 1999
Contact: McCarter JP
Unpublished (1999)
The Washington Univ. Nematode EST Project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
The library was constructed by Dr. Thomas Nutman and colleagues of
NIAID, NIH (tnutman@nih.gov). DNA Sequencing by: Washington
University Genome Sequencing Center St. Louis.
High quality sequence stop: 384.
FEATURES
source
location/Qualifiers
1..407
/organism="Strongyloides stercoralis"
/strain="Rhabditiform larvae obtained from gerbils"
/db_xref="taxon:6248"
/clone_lib="TBN95TM-SSR"
/lab_host="XL-1 Blue MRF" (Stratagene)"
/notes="Vector: Lambda Uni-ZAP XR (Stratagene); Site_1:
EcoRI; Site_2: XhoI; mRNA was purified from 2 x 10E3
rhabditiform larvae which had been isolated from gerbils
experimentally infected with larvae originally isolated
from experimentally infected dogs. cDNA was constructed
and, using adaptors, was cloned unidirectionally into the
vector from the EcoRI site to the XhoI site. The library
has an unamplified titer of 1 x 10E5 pfu/ml and an
amplified, undiluted titer of 9 x 10E11 pfu/ml. The
average insert size of the unamplified library is 675 bp
(range, 100-1700)."
BASE COUNT 152 a 63 c 74 g 118 t
ORIGIN

Query Match 79.1%; Score 17.4; DB 110; Length 407;
Best Local Similarity 94.7%; Pred. No. 1.8e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 ttgatcacaaagatttgata 22
|||||
Db 334 TTGATCACAGATTGTGATA 316

RESULT 12
AQ497779 529 bp DNA GSS 28-APR-1999
LOCUS HS_5066_B1_E05_17A RPCI-11 Human Male BAC Library Homo sapiens
DEFINITION genomic clone Plate-642 Col-9 Row-J, DNA sequence.
ACCESSION AQ497779
VERSION AQ497779.1 GI:4697902
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 529)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallaceu.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Resear h Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 642 row: J column: 9
Seq primer: T7
Class: BAC ends
High quality sequence stop: 529.
FEATURES
source
location/Qualifiers
1..529
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate-642 Col-9 Row-J"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/notes="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methyase. Size selected DNA was cloned into the
pBACE3.6 vector at EcoRI sites"
BASE COUNT 145 a 99 c 109 g 165 t
ORIGIN

Query Match 79.1%; Score 17.4; DB 158; Length 529;
Best Local Similarity 94.7%; Pred. No. 1.8e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 ttggatcacaaagatttgat 21
|||||
Db 403 TTGGTCACAGATTTGAT 421

RESULT 13
AW695022/c 674 bp mRNA EST 15-JUN-2000
LOCUS NF082E04ST1F1034 Developing stem Medicago truncatula cDNA clone
DEFINITION NF082E04ST 5', mRNA sequence.
ACCESSION AW695022
VERSION AW695022
KEYWORDS EST.
SOURCE barrel medic.
ORGANISM Medicago truncatula

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
Fabales; Fabaceae; Papilionoideae; Medicago.
1 (bases 1 to 674)
He,X.-Z., Shadle,G., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell
,C.J., Flores,H.R., Inman,J.T., Weller,J.W., May,G.D. and Dixon
,R.A.
Expressed Sequence Tags from the Samuel Roberts Noble Foundation
Medicago truncatula stem library
Unpublished (2000)
Contact: Dixon RA

```

Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel: 580 221 7302
Fax: 580 221 7380
Email: radixon@noble.org

Insert Length: 674 Std Error: 0.00
Plate: 082 row: E column: 04
Seq primer: TCACACAGAAACAGCTATGAC.
Location/Qualifiers

FEATURES

source

1. .674
/organism="Medicago truncatula"
/db_xref="taxon:3880"
/clone="NF082E04ST"
/tissue_type="stem"
/dev_stage="pooled developmental"
/note="Vector: Lambda Zap; Contains a mixture of
internodal stem segments"

BASE COUNT 189 a 149 c 133 g 203 t

ORIGIN

Query Match 79.1%; Score 17.4; DB 94; Length 674;
Best Local Similarity 94.7%; Pred. No. 1.9e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 ttggatcacagaatttgat 21

||||| |||||

Db 256 TTGGTCACTAGATTGAT 238

RESULT 14

BE283043

LOCUS

DEFINITION BE283043 762 bp mRNA EST 13-JUL-2000
601101323F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:3493829 5',
mRNA sequence.

ACCESSION

BE283043

VERSION

BE283043.1

KEYWORDS

EST.

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 762)
NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

Plate: LIAM8541 row: n column: 06
High quality sequence stop: 661.
Location/Qualifiers

FEATURES

source

1. .762
/organism="Mus musculus"
/strain="CZECH II (fetal)"
/db_xref="taxon:10090"
/clone="IMAGE:3493829"
/tissue_type="spontaneous tumor, metastatic to mammary."
Stem cell origin.
/lab_host="DH10B"
/note="Organ: lung; Vector: pCMV-SPORT6; Site: 1: SalI;
Site: 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigator

providing samples: Gilbert Smith, NIH

BASE COUNT 193 a 172 c 207 g 190 t

ORIGIN

Query Match 79.1%; Score 17.4; DB 106; Length 762;
Best Local Similarity 94.7%; Pred. No. 1.9e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 tgttgatcacagaatttg 19

||||| |||||

Db 692 TGTGGATCACAGACTTG 710

RESULT 15

BE642809

LOCUS

DEFINITION BE642809 837 bp mRNA EST 01-SEP-2000
Cri2_7_B04_SP6 Ceratopteris Spore Library Ceratopteris richardii
cDNA clone Cri2_7_B04 5', mRNA sequence.

ACCESSION

BE642809

VERSION

BE642809.1

KEYWORDS

EST.

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Ceratopteris richardii.
Ceratomyx richardii.
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Filicophyta;
Filicopsida; Filicales; Pteridaceae; Ceratopteris.
1 (bases 1 to 837)
Chatterjee.A., San Miguel,P., Stout,S.C., Banks,J. and Roux,S.J.
Expressed sequence tags of cDNA clones from a C. richardii library
Unpublished (2000)
Contact: Roux SJ
Section of Molecular Cell and Developmental Biology
University of Texas
Biology Building, Room 16, Austin, TX 78712, USA
Tel: 512 471 4238
Fax: 512 232 3402
Email: sroux@uts.cc.utexas.edu
Plate: Cri2_7 row: B column: 04
Seq primer: SP6.
Location/Qualifiers

FEATURES
source

1. .837
/organism="Ceratopteris richardii"
/cultivar="Brogn"
/db_xref="taxon:49495"
/clone="Cri2_7_B04"
/tissue_type="Ceratopteris Spore Library"
/cell_type="Spore"
/dev_stage="20 hours after germination initiation"
/note="Vector: pCMVSPORT6; EST sequence from cDNA library.
cDNA library constructed from mRNA isolated from C.
richardii spores that had developed for 20 hours after
their germination had been initiated by white light."
BASE COUNT 233 a 158 c 214 g 232 t

ORIGIN

Query Match 79.1%; Score 17.4; DB 110; Length 837;
Best Local Similarity 94.7%; Pred. No. 1.9e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 tgttgatcacagaatttg 19

||||| |||||

Db 332 TGTAGGATCACAGATTG 350

Search completed: March 27, 2001, 07:38:20
Job time: 4583 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 27, 2001, 06:54:38 ; Search time 207.51 Seconds
(without alignments)
2673.865 Million cell updates/sec

Title: us-09-380-826A-1

Perfect score: 1477
Sequence: 1 gatcatggctgaactaac.....ccgtaaatcattcttcgag 1477

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_36:*

```
1: /cgn2_2/gcgdata/geneseq/geneseq/NA1980.DAT:*
2: /cgn2_2/gcgdata/geneseq/geneseq/NA1981.DAT:*
3: /cgn2_2/gcgdata/geneseq/geneseq/NA1982.DAT:*
4: /cgn2_2/gcgdata/geneseq/geneseq/NA1983.DAT:*
5: /cgn2_2/gcgdata/geneseq/geneseq/NA1984.DAT:*
6: /cgn2_2/gcgdata/geneseq/geneseq/NA1985.DAT:*
7: /cgn2_2/gcgdata/geneseq/geneseq/NA1986.DAT:*
8: /cgn2_2/gcgdata/geneseq/geneseq/NA1987.DAT:*
9: /cgn2_2/gcgdata/geneseq/geneseq/NA1988.DAT:*
10: /cgn2_2/gcgdata/geneseq/geneseq/NA1989.DAT:*
11: /cgn2_2/gcgdata/geneseq/geneseq/NA1990.DAT:*
12: /cgn2_2/gcgdata/geneseq/geneseq/NA1991.DAT:*
13: /cgn2_2/gcgdata/geneseq/geneseq/NA1992.DAT:*
14: /cgn2_2/gcgdata/geneseq/geneseq/NA1993.DAT:*
15: /cgn2_2/gcgdata/geneseq/geneseq/NA1994.DAT:*
16: /cgn2_2/gcgdata/geneseq/geneseq/NA1995.DAT:*
17: /cgn2_2/gcgdata/geneseq/geneseq/NA1996.DAT:*
18: /cgn2_2/gcgdata/geneseq/geneseq/NA1997.DAT:*
19: /cgn2_2/gcgdata/geneseq/geneseq/NA1998.DAT:*
20: /cgn2_2/gcgdata/geneseq/geneseq/NA1999.DAT:*
21: /cgn2_2/gcgdata/geneseq/geneseq/NA2000.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1477	100.0	1477	19	V58896
2	814.6	55.2	1535	20	X83569
3	813.4	55.1	1512	20	X82000
4	811.6	54.9	1529	20	X83564
5	810.2	54.9	1508	20	Z31557
6	810.2	54.9	1508	21	Z45352
7	810	54.8	1529	20	X83567
8	808.4	54.7	1529	20	X83565
9	808.4	54.7	1529	20	X83566
10	807.8	54.7	1529	20	X81999
11	805.4	54.5	1521	16	Q92879
12	805.4	54.5	1526	16	T01866

13	805	54.5	1528	20	X83570
14	805	54.5	1535	20	X83568
15	803.8	54.4	1536	17	T18765
16	802.8	54.4	1477	20	Z10000
17	801.8	54.3	1512	20	X81996
18	801.6	54.3	1555	17	T29142
19	801.6	54.3	1555	17	T24294
20	801.4	54.3	1450	21	Z45351
21	800.4	54.2	1532	13	Q26729
22	796.4	53.9	1506	20	Z31560
23	795.6	53.9	1532	15	Q64008
24	794.8	53.8	1446	17	T10955
25	794.8	53.8	1535	19	V24295
26	793.6	53.7	1535	20	X84802
27	791.8	53.6	1477	20	X26285
28	789.2	53.4	1452	20	X82004
29	789	53.4	1535	20	X83571
30	786.8	53.3	1477	17	T43670
31	785.8	53.2	1455	20	X77384
32	784.4	53.1	1566	17	T18645
33	781	52.9	1540	10	N91514
34	780.2	52.8	1542	17	T18759
35	780	52.8	1542	17	T29140
36	780	52.8	5098	20	X24984
37	780	52.8	5341	20	X24986
38	779.8	52.8	1474	17	T18766
39	779.4	52.8	1548	20	X84804
40	778.4	52.7	1556	20	X81995
41	778.4	52.7	5097	20	X24983
42	777.2	52.6	5105	20	X24989
43	776.8	52.6	5014	20	X24987
44	775.6	52.5	1513	20	X82003
45	775.4	52.5	1450	20	X82005

C

ALIGNMENTS

RESULT 1

ID V58896 standard; DNA: 1477 BP.

XX V58896;

AC 20-JAN-1999 (first entry)

DT L. fainei nucleotide sequence.

DE Infection; pathogenic Leptospira; protective immunity; therapy;
KW diagnosis; ss.

OS Leptospira fainei.

PN WO9840099-A1.

PD 17-SEP-1998.

PF 06-MAR-1998; 98WO-AU00145.

PR 07-MAR-1997; 97AU-0005494.

XX (AGRI-) AGRIC VICTORIA SERVICES PTY LTD.

XX (PIGR-) PIG RES & DEV CORP.

PI Chappel RJ;

XX WPI; 1998-520791/44.

XX New isolated pathogenic Leptospira bacterium - useful for, e.g
PT developing products for conferring protective immunity, and for
PT prophylactic or therapeutic treatment

XX Claim 15; Page 69-70; 94pp; English.

P.cepacia 16S rRNA

XX This sequence represents a *Leptospira* DNA sequence isolated from the
 CC pathogenic *Leptospira* (LS) bacterium of the invention. The bacterium
 CC belongs to serogroup Hurstbridge or serovar hurstbridge or the species
 CC *L. falnet*. The LS bacterium can be used for conferring protective
 CC immunity against pathogenic LS bacteria in humans or animals. The
 CC bacteria can also be used for prophylactic or therapeutic treatment of LS
 CC infections. The DNAs and antibodies may also be used for detection and
 CC diagnosis of past or present LS infection.
 XX

XX Sequence 1477 BP; 390 A; 334 C; 439 G; 314 T; 0 other;
 XX

Query Match 100.0%; Score 1477; DB 19; Length 1477;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1477; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gatcatggtcagaactaactgctggcgccgctcttaaacatgcaagtcgagcggttag 60
 DB 1 gatcatggtcagaactaactgctggcgccgctcttaaacatgcaagtcgagcggttag 60
 QY 61 caatacctagcggcgaaacgggtgagtaacacgctgtaattcttccctccgagtcgggataa 120
 DB 61 caatacctagcggcgaaacgggtgagtaacacgctgtaattcttccctccgagtcgggataa 120
 QY 121 ctttcgaaagaaagctaaatcccgatagtccttgcgacacgctgtaattcttccctccgagtcgggataa 180
 DB 121 ctttcgaaagaaagctaaatcccgatagtccttgcgacacgctgtaattcttccctccgagtcgggataa 180
 QY 181 agatttattgcttgagatgaaccccgccgagcttagcttagtgtaggtaattgctca 240
 DB 181 agatttattgcttgagatgaaccccgccgagcttagcttagtgtaggtaattgctca 240
 QY 241 ccaagcgacgacgctgtagccgctgagaggtgtcccgccacaaatgaactgagacac 300
 DB 241 ccaagcgacgacgctgtagccgctgagaggtgtcccgccacaaatgaactgagacac 300
 QY 301 ggtcactactcactcagggagcagcagtaagaattcttgcctcaatgggggaaacccctgaa 360
 DB 301 ggtcactactcactcagggagcagcagtaagaattcttgcctcaatgggggaaacccctgaa 360
 QY 361 gcagcgacgacgctgtagcaggaaggtcttcgagatgtaaaatcattagcgagaaaa 420
 DB 361 gcagcgacgacgctgtagcaggaaggtcttcgagatgtaaaatcattagcgagaaaa 420
 QY 421 ataagcagaatgtgatgtgtagcttccctcctaaagcagcagctaacctacgtgcccagcagc 480
 DB 421 ataagcagaatgtgatgtgtagcttccctcctaaagcagcagctaacctacgtgcccagcagc 480
 QY 481 cgcggtatacgtatgtgtagcagcgttcttcggaatcattggcgtaaaagggtgcgtagg 540
 DB 481 cgcggtatacgtatgtgtagcagcgttcttcggaatcattggcgtaaaagggtgcgtagg 540
 QY 541 cggatttgaatcaggtgtgaaacgctgggctcaaccctgctgacttgaacta 600
 DB 541 cggatttgaatcaggtgtgaaacgctgggctcaaccctgctgacttgaacta 600
 QY 601 caagctcggagtttggagaggaagtggaattccaggtgtagcgtggaatcgttagat 660
 DB 601 caagctcggagtttggagaggaagtggaattccaggtgtagcgtggaatcgttagat 660
 QY 661 atctggaggaacacacgctggcgaagcagcttctggtcacaactgaagcgtgagcagc 720
 DB 661 atctggaggaacacacgctggcgaagcagcttctggtcacaactgaagcgtgagcagc 720
 QY 721 aaagcgtgggttagtaacgggattagatacccggttaattccacgcccctaaacgttgccta 780
 DB 721 aaagcgtgggttagtaacgggattagatacccggttaattccacgcccctaaacgttgccta 780
 QY 781 ccagttgttgggggttttaacctcgtgaacacctaagcgttagtaagtagaccgctgg 840
 DB 781 ccagttgttgggggttttaacctcgtgaacacctaagcgttagtaagtagaccgctgg 840

QY 841 ggaactatgctcgaagagtgaaactcaaggaattgacgggggtccgcacaaagcgtgga 900
 DB 841 ggaactatgctcgaagagtgaaactcaaggaattgacgggggtccgcacaaagcgtgga 900
 QY 901 gcatgtggttatttcgatgatacccccaaaacccctcactggcttgacatggatctgaa 960
 DB 901 gcatgtggttatttcgatgatacccccaaaacccctcactggcttgacatggatctgaa 960
 QY 961 tcatgtagagatatatgagccttcgggcagattcacaggtgctgcattgtgtcgtcagc 1020
 DB 961 tcatgtagagatatatgagccttcgggcagattcacaggtgctgcattgtgtcgtcagc 1020
 QY 1021 tcatgtcgtgagattggtggttaagtcgccgaacgagcgaacccctatcgtatgttgc 1080
 DB 1021 tcatgtcgtgagattggtggttaagtcgccgaacgagcgaacccctatcgtatgttgc 1080
 QY 1081 accttaagtgtggcactggtacgaaactgcggtgacaaaacccgaggaagcgggatga 1140
 DB 1081 accttaagtgtggcactggtacgaaactgcggtgacaaaacccgaggaagcgggatga 1140
 QY 1141 cgtcaaatccctcagtcgcttattatccagggccacacacgtgctacaatggccgatacag 1200
 DB 1141 cgtcaaatccctcagtcgcttattatccagggccacacacgtgctacaatggccgatacag 1200
 QY 1201 aggtcgcgaactgcgaagggaggttaattcttcaaaagtcggtccagttcggattggg 1260
 DB 1201 aggtcgcgaactgcgaagggaggttaattcttcaaaagtcggtccagttcggattggg 1260
 QY 1261 gctcgaactgcgaacccatgaatgcggaatcgctagtaatcgcggtacagtcgcgcgg 1320
 DB 1261 gctcgaactgcgaacccatgaatgcggaatcgctagtaatcgcggtacagtcgcgcgg 1320
 QY 1321 tgaatacgttcccgacgttgcacaccccgctcacacacacacacacacacacacacac 1380
 DB 1321 tgaatacgttcccgacgttgcacaccccgctcacacacacacacacacacacacacac 1380
 QY 1381 gaagtgctcgaacaggtacacgttaaatcgattcctcgcag 1477
 DB 1441 gaagtgctcgaacaggtacacgttaaatcgattcctcgcag 1477
 DB 1441 gaagtgctcgaacaggtacacgttaaatcgattcctcgcag 1477

RESULT 2

XX X83569
 ID X83569 standard; DNA; 1535 BP.
 XX
 AC X83569;
 XX
 DT 21-DEC-1999 (first entry)
 XX
 DE 16S rDNA gene fragment from marine bacterium, isolate K3-3.
 XX
 KW Monitoring; oil; contamination; sea water; detection; flagellum;
 KW Gram-negative bacterium; Proteobacteria; glucose; carbon source;
 KW alkane; 16S rDNA gene; ds.
 XX
 OS Proteobacteria.
 XX
 PN JP11243967-A.
 XX
 PD 14-SEP-1999.
 XX
 PF 04-MAR-1998; 98JP-0069399.
 XX
 PR 04-MAR-1998; 98JP-0069399.
 XX
 PA (KAIY-) KAIYO BIOTECHNOLOGY KENKYUSHO KK.
 XX
 DR WPI; 1999-564435/48.
 XX


```
QY 690 cttgtgctcaaaactgacgtgagcgacgaaacgctgggtagtaaaacgggattagata 749
Db 720 cctcctgactgacattgacgtgagtgctgaaacgctggggagcaaacagattagata 779
QY 750 ccccggttaacacccctcaacgctgtctaccagttgttgggggtttta--accctcag 807
Db 780 cccgtgtagtcacacgcgttaaacagatgctctactagctgttgggaatcttagtatcttgg 839
QY 808 taacgaacctaacgattgaatagaccctgctgggactatgctgcgaagagtgaaactca 867
Db 840 tgacgaagttaacgcgaataagtagaccgctggtggagtagccgcgaaggtttaaactca 899
QY 868 aaggaattgacggggtccgcacaacgagcgtggagcatgtgtttaatcgtatgatacccc 927
Db 900 atgtaattgacggggccgcacacgagcgtggagcatgtgtttaatcgtatgatacccc 959
QY 928 aaaaacctcaacctggcttgacatggatctga-atcatgtagatagatatagaccttcgg 986
Db 960 aagaaccttaccagccttgacatccttggaaacttctagagatagattggtgccttcgg 1019
QY 987 --gcagattcacagtgctgcatggtgtctgctacgtcgtgtcgtgagatgttgggttaa 1044
Db 1020 gagccaagtgacaggtgctgcatggtgtcgtcagctcgtgtcgtgagatgttgggttaa 1079
QY 1045 gtcccgcaacgagcgaacccctatcgtatgttg--taccttaagttggcactgata 1101
Db 1080 gtcccgtaacgagcgaaccccttgccttagttgcccacacataatgttgggaactctag 1139
QY 1102 cgaactccggtgacaaacggaggaagcggtgagatgacgtcaaatctctcatgacctt 1161
Db 1140 ggagactccggtgacaaacggaggaaggtgggacacgctcaagatcatcatgccctt 1199
QY 1162 atgtccagggccacacgctgtcataatggccgatacacagaggggtgcgaactcgcgaag 1221
Db 1200 acggcctgggtcacacagctgtcataatggcgagcacagagggcgaggtcgcgaggc 1259
QY 1222 ggagctaatcttaaaagtcgtccagttcgagttgggttgcgaactcgaacccatga 1281
Db 1260 caagcaatcccttaaaactgtctgtagtccggatgtagtgcgaactcgaactcga 1319
QY 1282 agtcggaatcgtagtaatacgcggtacgacatgcgcggtgtaatacgttcccgaccttg 1341
Db 1320 agtcggaatcgtagtaatacgcggtacgagaaatgcgcggtgtaatacgttcccgacctg 1379
QY 1342 tacacacgcgcgtcacacacacactgagtggggagcacccgaaagtgttcttaacagt 1401
Db 1380 tacacacgcgcgtcacacacacactgagtgaggatggattgcaccagaaagtggtg-atagtctaacctt 1438
QY 1402 aaggagacagactactaagtgaaactcgtaaaggggtgaagtcgtcaacaaggtaac 1459
Db 1439 cgggaggacgttcacacacggtgtgttcatgactgggtgaaagtcgtcaacaaggtagc 1496
```

RESULT 8

X83565

ID X83565 standard; DNA; 1529 BP.

XX AC

XX AC

XX AC

XX 21-DEC-1999 (first entry)

XX DE 16S rDNA gene fragment from marine bacterium isolate wf-1.

XX KW

KW Monitoring; oil; contamination; sea water; detection; flagellum;

KW Gram-negative bacterium; Proteobacteria; glucose; carbon source;

KW alkane; 16S rDNA gene; ds.

XX OS

XX Proteobacteria.

XX PN

XX JPI1243967-A.

XX PD

XX 14-SEP-1999.

XX

XX

PF 04-MAR-1998; 98JP-0069399.

XX PR 04-MAR-1998; 98JP-0069399.

XX PA (KAIY-) KAIYO BIOTECHNOLOGY KENKYUSHO KK.

XX DR WPI; 1999-564435/48.

XX PT Monitoring of oil contamination of sea water - where oil contamination

PT is evaluated by detection of a microbe having properties from e.g.

PT having no flagellum, being a Gram-negative bacterium, belonging to

PT Proteobacteria, gamma subdivision, etc.z

XX Claim 3; Page 5; 15pp; Japanese.

XX The invention relates to a method for monitoring oil contamination of

CC sea water by detecting, in the sea water, a microbe having the following

CC properties: (1) it has no flagellum; (2) it is a Gram-negative bacterium;

CC (3) it belongs to Proteobacteria, gamma subdivision; (4) it cannot

CC assimilate glucose as a single carbon source; and (5) it efficiently

CC assimilates at least one of 10-30C n-alkanes. This sequence represents

CC a fragment of the 16S rDNA gene from the microbe of the invention,

XX isolate wf-1.

SQ Sequence 1529 BP; 376 A; 352 C; 489 G; 312 T; 0 other;

Query Match 54.7%; Score 808.4; DB 20; Length 1529;

Best Local Similarity 76.1%; Pred. No. 1.3e-241;

Matches 1140; Conservative 0; Mismatches 316; Indels 42; Gaps 10;

QY 1 gatcatggtcctgagaactaacgctgcgcggcgtctttaaactgcaagtcgagcggggtag 60

Db 2 gatctggtcctgagattgaacgctgcgcggcgtctttaaactgcaagtcgagcggggtag 61

QY 61 caatac-----ctagcgcgcgaacgggtgagtaaacacgctggaattcttc 104

Db 62 atcctagcttgtagagagggcgtcgcgcggcgtctttaaactgcaagtcggaattcttc 121

QY 105 tcgcagctggtgataacttccgaaaggaaagcttaataccggaatgctctgttggtatcac 164

Db 122 cattagtggggataaacctcgggaaaccagcgttaataccggaatgctctgttggtatcac 181

QY 165 aagatttgataggttaaaagtattatgcttgagagatgagccgcgcgattagctagctg 224

Db 182 agca--gggacaccttgcgccttctgtgatgtagtgcgtgcgtgcgtgcgtgcgtgcgtg 239

QY 225 gtgagtaagtgcctcaccaagcgcagatcgtagccgcctcgaaggggtgcgcgcac 284

Db 240 gtgagtaagtgcctcaccaagcgcagatcgtagccgcctcgaaggggtgcgcgcac 299

QY 285 aatggaactgagacacggtccatactcctcacggagcagcaggttaagaatcttgcctcaa 344

Db 300 accggagctgagacacggtccatactcctcacggagcagcaggttaagaatcttgcctcaa 359

QY 345 tggggaaacacctgaagcagcagcgcgcgtgaaacgaaaggtcttcggattgtaag- 403

Db 360 tggggaaacacctgaagcagcagcgcgcgtgaaacgaaaggtcttcggattgtaag- 419

QY 404 ---ttcattagcagcagaaataaagcagcaatgtg-----atgatggctaccctgcct 451

Db 420 atttcagtagggaggaaggtcttattccttaacacggaatgtagtactgagcttaccctcacg 479

QY 452 a--aagcaccgcttaactacgtgccagcgcgcgtgaaacgaaaggtcttcggattgtaag- 509

Db 480 aagaagcaccgcttaactacgtgccagcgcgcgtgaaacgaaaggtcttcggattgtaag- 539

QY 510 tcggaatcatctgggcgttaaaaggtgcgtgagcaggtttgtaagtcaggtgtgaaactgc 569

Db 540 tcggaatcatctgggcgttaaaaggtgcgtgagcaggtttgtaagtcaggtgtgaaactgc 599

QY 570 gggctcaacccctggcctgacattgaaactacaagctgaggtttgggagaggaaggtg 629

QY 510 tcggaaatcattggcggttaaaaggttcgtagcggtgatttctgaagtccaggtgtgaaactgc 569
 Db 540 tcggaaatcattggcggttaaaaggttcgtagcggttctgaagtccaggtgtgaaagcccc 599
 QY 570 gggctcaacccgtgcccctgacactgaaactcaaaactcggagtttggagagcgcaagtgg 629
 Db 600 gggctcaacccgtgaaactgcaatttgaactgcaagctgagatgcaatgagaggggtgg 659
 QY 630 aattccaggttagcgggtgaaatcgctagatatcttggaggaaacacacagtgccgaaggcga 689
 Db 660 aatttccgggttagcgggtgaaatcgctagatatcttggaggaaacacacagtgccgaaggcgg 719
 QY 690 cttgctggctcaaaactgacgtgagcgcaagaaacgctgggttagtaaaacgggattagata 749
 Db 720 cctcctgactgacattgacgtgagcggtgcaaaacgctggggagcaaacagagattagata 779
 QY 750 ccccggttaacccacccctcaacgtgtctacacagttgttgggggtttta--acccctcag 807
 Db 780 cccgtgtagtccacccgtcaaacgattgctactactgcttgggaattcttagtattcttgg 839
 QY 808 taacgaacctaaacgattaaagttagaccgctggtggactatgctcgaagagtgaactca 867
 Db 840 tgacgaagttaacgcgataaagttagaccgctggtggagtagcagcgcaaggttaaaactca 899
 QY 868 aaggaaatgacgggggtccgcacacagcggtggagcatgtgtttaaattcgatgatacccc 927
 Db 900 aatgaattgacgggggtccgcacacagcggtggagcatgtgtttaaattcgatgatacccc 959
 QY 928 aaaaacctcacctggcctgacatggatctga--atcaltagagatatatagaccttcgg 986
 Db 960 aagaaccttaccagccttgacatccttggaaacttctagagatgagttggccttcgg 1019
 QY 987 --gcagattcacagtgctgcaatgttgcctcagctcgtgctggaatgttgggttaa 1044
 Db 1020 gagcaagtgcagagtgctgcaatgttgcctcagctcgtgctggaatgttgggttaa 1079
 QY 1045 gtccgcgaacgagcgaacccctatcgtattgtgc---taccttaagttgggcaactgta 1101
 Db 1080 gtccgcgaacgagcgaaccccttgccttagttgccagcacataatgttgggaactctag 1139
 QY 1102 cgaactgcggtgacaaacggaggaagcggtgagtagcgtcaaatccctcctatggccttt 1161
 Db 1140 gagactgcggtgacaaacggaggaaggttgggacagctcaagtcacatcctatggccttt 1199
 QY 1162 atgtccagggccacacagtgctacaatgtgcccgaatcacagaggggtgcgaactcgcgaag 1221
 Db 1200 acggcctgggtcacacagtgctacaatgtgcccgaatcacagaggggtgcgagagc 1259
 QY 1222 ggagtaattcttaaaagtcgctccaggttcgagtagtgggttcgcaactcgaccctatga 1281
 Db 1260 caagcaaatcccttaaaactgtctgtagtccggattggagttcgaactcgactccatga 1319
 QY 1282 agtcgaaatcgctagtaatcgcggaatcgacatgcccgggtgaaatcgttcccggaccttg 1341
 Db 1320 artcgaatcgtagtaatcgcggaatcgagaatgcgcggtgaaatcgttcccggaccttg 1379
 QY 1342 tacacacgcccgttcacacacacactgagtggggagcaccgcaagtggttttgaacogt 1401
 Db 1380 tacacacgcccgttcacacacacactgagtggtgattgacacagaagtgg--atagctaaacct 1438
 QY 1402 aaggagacagactaactaaggtgaaactcgttaaaaggggtgaaagtgcgaacaagggtacc 1459
 Db 1439 cgggaggacgttcaacacgggtgtgttcaatgactgggttgaagtgcgaacaagggtacc 1496

RESULT 10

X81999
 ID X81999 standard; DNA; 1513 BP.

XX
 AC X81999;

XX
 DT 10-SEP-1999 (first entry)

XX

DE B. cereus BCM 4 partial 16S rRNA gene sequence.
 XX Taxane; pacilitaxel; plant; bacteria; mutation; cancer; 16S rRNA; ss.
 OS Bacillus cereus.
 XX WO9932651-A1.
 XX 01-JUL-1999.
 XX PD 18-DEC-1998; 98WO-CA01150.
 XX PF 22-DEC-1997; 97US-0995960.
 XX PR (BCMB-) BCM DEV INC.
 XX PI Bolssinot M, Gagne M, Harvey M, Helie M, Landry N;
 XX Page M;
 XX WPI; 1999-418940/35.
 XX DR Production of taxane and pacilitaxel compounds
 XX PS Disclosure; Fig 8E; 93pp; English.
 XX The invention relates to a new method for the production of taxane and
 CC pacilitaxel compounds that comprises culturing bacteria isolated from a
 CC plant species of Taxus or bacteria produced by mutating the isolated
 CC bacteria. The methods can be used for the production of taxanes such as
 CC pacilitaxel, 10-deacetylcephalomanine, 7-epitaxol, 10-deacetyl-7-epi-
 CC taxol, 7-epicephalomanine, baccatin III, 10-deacetyl-baccatin III,
 CC cephalomanine, 7-epibaccatin III, 7-xylosyltaxol, 7-xylosyl-
 CC cephalomanine, taxagifine, delta-benzoyloxy taxagifine, 9-acetyloxy
 CC taxusin, 9-hydroxy taxusin, taxane Ia, taxane Ib, taxane Ic, or taxane
 CC Id (claimed). The taxane and pacilitaxel compounds can be used for the
 CC treatment of cancers. The methods can provide for the production of
 CC taxanes and pacilitaxel at a concentration of 1-25 mu g/l.
 XX Sequence 1513 BP; 387 A; 341 C; 467 G; 316 T; 2 other;

Query Match 54.7%; Score 807.8; DB 20; Length 1513;
 Best Local Similarity 75.1%; Pred. No. 2.1e-241;
 Matches 1130; Conservative 1; Mismatches 328; Indels 46; Gaps 8;

QY 1 gatcctggctcagaactaacgctggtggtggtggtttaaactgacgagcg-----54
 Db 8 gatcctggctcagaactaacgctggtggtggtttaaactgacgagcgaggaatgga 67
 QY 55 -----gggtgagcaatacctagcggcgacggtgagtaaacacgt-ggtaacttc 103
 Db 68 ttgagagctgtctcagaagtagcggcgacggtgagtaaacacgtgggtaacctgc 127
 QY 104 ctccagctggtgagtaaaccttccgaagaaagcaagtaaacacgata------gtcctgt 156
 Db 128 ccataagcgtggtgagtaaaccttccgaagaaagcaagtaaacacgataattttgacatgc 187
 QY 157 tggatcacagatttgatgagtaaacatttattgcttggagatgacccggtgacgattta 216
 Db 188 atggttcgaattgaaaggcgttccgtgctcacttattggtgacccggtcgcgattta 247
 QY 217 gctagttggtgaggttaattggtccacaaagcagcagctgtagccggtcgtgagaggtgt 276
 Db 248 gctagttggtgaggttaacggtcaccacaaagcagcagctgtagccggtcgtgagaggtga 307
 QY 277 ccggccacaatggaactgagacacggtccatactctacgggagggcagcagtttaagaatc 336
 Db 308 tcggccacactgggactgagacacggtccacagctcctacgggagggcagcagtagggaatc 367
 QY 337 ttgctcaatgggggaaacccctgaagcagcagcgcggtgaaacgaagaggttcttcggat 396
 Db 368 ttccgcaatggacgaagctgacgagcaacgcccgcgtgagtgatgagaggttcttcgggt 427

QY	397	tgtaaagtctcattgacgaggaataaagcagcaatdgtatgatagttactcgtcct-----	451
Db	428	cgtaaactctgtttagggaaagaacaagtgtctagtgtgaataagctggcaccttgacgg	487
QY	452	-----aaacacccgcttaactactgtgccagcagccgcggtaaatcgtatgtgtgc	500
Db	488	taactaaccaagaagccacccgcttaactactgtgccagcagccgcggtaaatcgtatgtgtgc	547
QY	501	aagcgtgttcggaatcatttggcgttaaaaggtgctgtagcggaatttgaatcgaagttgt	560
Db	548	aagcgttaaccggaattatttggcgttaaaagcgcgcgaggtgttcttaggtctctgatgt	607
QY	561	gaaaaactgcgggctcaacccttgccctcacttbaaaactcaaaactgaagtttgggaaga	620
Db	608	gaaagccacccggtcaaaccttgagggttcatttgaacctgggagacttgagtgcagaaga	667
QY	621	ggcaagtggaaattccaggtgtagcgtgaaatcgctagatatctcggaggaaacaccagttgg	680
Db	668	gaaagtggaaattccatgtgtagcgtgaaatcgctagagatatggaggaaacaccagttgg	727
QY	681	cgaagcgacttgcctgcctcaaaactgcgcctgtaggcagcaaaaagcgtgggtagtaaacagg	740
Db	728	cgaagcgacttcttgtctgttaactgacatcgtaggcgcgaaagcgtggggagcaaacagg	787
QY	741	gattagatacccggttaataccacgccttaaacgcttgtctacagatttcttgggggtttt--	798
Db	788	gattagataccctgtgtagtcacgcgcgttaaacgatgagtctaagttgtatagagggtttcc	847
QY	799	aaocctcagtaacgaacctaaacggttaagttagaccgcctcgtgggacttatctcgcagaag	858
Db	848	gcccttttagtgtgaagttaaacgcattaaagcactcgcctcgtgggagtcacgcccgaagc	907
QY	859	tgaacctcaagaattatgcgggggtccgcacaacgcgtggagcatgtggtttaaattcga	918
Db	908	tgaacctcaagaattatgcacggggccgcacaacgcgtggagcatgtggtttaaattcga	967
QY	919	tgatcccccaaaacctcacttggccttgaocat--ggatctgaaatcatgtagagatatgt	977
Db	968	agcaacgcgaagaaccttaccagggtcttgacatcctctgaaaaaccttagagataggcctt	1027
QY	978	agccttcgg--gcagattcacaggtgtcgtatgtgttcgtcagctcgtgtcgtcgagatg	1035
Db	1028	ctccttcgggagcagagtgacaggtgggtcgtatgtgtgttcgtcagctcgtgtcgtcgagatg	1087
QY	1036	tgggttaagtccccgcacaacgcagcgaacccct--atcgtatgttgcctaccttaagttgggc	1094
Db	1088	tgggttaagttcccgcaacgcagcgaaccccttgatctcttagttgccatcattaaattgggc	1147
QY	1095	actggtaacgaacctgcggttgacaaaccggaaggaagcggggatgaagctcaaatccctcat	1154
Db	1148	acttaagtgtagtgcggttgacaaccggaaggaaggtggggagacgtcgaatcgaatcgaat	1207
QY	1155	ggccttatgttccagggccacacagctgtaacaatggccgatcacagaggttcgcgaactc	1214
Db	1208	gycccttatgacctgggctacacagctgtaacaatggacggttacaaaagagctgcaagacc	1267
QY	1215	gcaagggagcgttaactctataaagtctgtcccagttcggatttgggggtctcgaactcgac	1274
Db	1268	gcgaggttgagagctaaactctcaaaaaccgttctcagttcgtatgttaggtctcaactcgcc	1327
QY	1275	ccccgaagtcgaatcgtctagtaactgcggtacagcatgcgcggttgaaatcagcttcccgc	1334
Db	1328	tacatgaagctggaaatcgtctagtaactgcgcggttcagcatgcgcggttgaaatcagcttcccgc	1387
QY	1335	gaccttgtacacacccgcccttcacacacctgagttggggagcaccgcgaagtggcttttgt	1394
Db	1388	ggccttgtacacacccgcccttcacacacctgagttggggagcaccgcgaagtggcttttgt	1447
QY	1395	taacggttaagagacagactactaaagttgaaactcgttaagggggtgagtcgtataacaaag	1454
Db	1448	aaccttttggagccagcccttaaggttgggacagatgatgttgggttgaaagtcgtataacaaag	1507
QY	1455	gtacc	1459

```
Db 323 cactgctcagactcctacggaagcagcagtagggaatacatccgcaatggcggaagcct 382
Qy 358 gaagcagcagccgcgtgaaacgaagaaggtcttcgagttgtaaaagttcattagcgagga 417
Db 383 gacggtgcaacccgcgtgaacgatgaaggttttcggtatcgtaaagtctgttatgaggg 442
Qy 418 aaaaataagcagaatgtagtgatgctgct-----aaagcaccgg 461
Db 443 agaaacaagtgccgttcgaataagtcgacaccttgacggtacctcaacagaaagcccccgg 502
Qy 462 ctaactacgtgccagcagccgcgttaatacgtatggtgcaagcgttcttcggaatcattg 521
Db 503 ctaactacgtgccagcagccgcgttaatacgtatggtgcaagcgttcttcggaatcattg 562
Qy 522 ggcgtaaagggctgtaggcggaattgttaagtcaggtgtgaaacactgcgggtcacaacccg 581
Db 563 ggcgtaaagcgcgcagcggctctcttaagtcctgatgtgaagcgcacgcgtcacaacgt 622
Qy 582 tggcctgcaactgaaactacaagctgagtttggaagcgaagtggaattccaggtgt 641
Db 623 ggaaggtcatgtgaaactgggggaacttgagtgtaggaggaagtggaattccacggtgt 682
Qy 642 agcgtgaaatcgtagatatctgaggaacaccagtggtggaagcagcttctgtgctca 701
Db 683 agcgtgaaatcgtagatatctgaggaacaccagtggtggaagcagcttctgtgctca 742
Qy 702 aaactgacgtgagcagcagaagcgtggttagtaaacgggtagatacccccggttaatcc 761
Db 743 caactgacgtgagcagcagaagcgtggttagtaaacgggtagatacccccggttagtcc 802
Qy 762 acgcctaaacgttctaccagttgttgggggttt--taaccctcagtaacgaaccta 819
Db 803 acgcgtaaacgtgagtgctaggtggttaggggttttccatacccttagtgcggaagttaa 862
Qy 820 cggattaaagtagaccgcctggggaactatgctgcgaagtagtgaactcaaaaggaattgacg 879
Db 863 cacattaaagcactcgcctggggaagtagcgcgcgaagtggaactcaaaaggaattgacg 922
Qy 880 ggggtccgcagaagcgggtggaacatggtgttaattcgatgatacccccaaaacccacc 939
Db 923 ggggcccgcagaagcagtgagcatggtgttaattcgaaacgcgaagacacattacc 982
Qy 940 tgggtctgaca-----tggatctgaatcatgtagagatatagccttcgggcagattc 994
Db 983 aggtcttgacatcctctgacacctctggagacagagcgttcccttcgggggacagagtg 1042
Qy 995 acaggtctgcattgtctgcagctcgttcgtgagatggttggttaagttcccgcaac 1054
Db 1043 acaggtggtgcattgtctgcagctcgttcgtgagatggttggttaagttcccgcaac 1102
Qy 1055 gaggcacaacctc-atcgatgttctaccctaaagttaggtgggaactggtgcgaactgcgcg 1113
Db 1103 gaggcacaaccttgatcttagttgcagcattcagttggtggaacttaaggtgactgcgg 1162
Qy 1114 tgacaacccgaggaagcgggagtagcgtcaaatccatcgatgcctttagtccaggacc 1173
Db 1163 tgataaacccggaaggttgggagtgacgtcaaaatcatatgcctttagtaccctgggct 1222
Qy 1174 acacacgtgctacaatggccgatacagaggggtcgcccaactcgcaagagggagctaatctc 1233
Db 1223 acacacgtgctacaatgggtgtacaaagggcagcgagacccgcgaggttaagcgaatccc 1282
Qy 1234 taaaagtcggtcccgatctcggaattgggtctgcacactcgaccccatgaagtcggaatcgc 1293
Db 1283 ataaagcattctcagttcggaattgcaggtcgcaactcgctgctgtagcgcgggaattgc 1342
Qy 1294 tagtaatcgcgatcagatgcgcgggtggaatcacgttcccgacactgtgacacacgcgc 1353
Db 1343 tagtaatcgcgatcagatgcgcgggtggaatcacgttcccgacactgtgacacacgcgc 1402
Qy 1354 gtcacacacactgagtggggagcaccacgaagtgggtctttgttaacctgaagcagacagac 1413
Db 1403 gtcacacacagaggttgttaacacccgaagtcg-gtgcggtaacctttgtgagcagacc 1461
```

```
Qy 1414 tactaagtgaaactcgttaaaaggggtgaagtcgtaacaaggtacc 1459
Db 1462 gncgaaggtggacacagatgattgggtgagtcgtaacaaggtatc 1507
```

RESULT 12

T01866

ID T01866 standard; DNA; 1526 BP.

XX

XX T01866;

XX

XX 03-AUG-1999 (first entry)

XX

XX P.cepacia 16S rRNA gene sequence.

XX

XX 16S rRNA; KK01; primer; PCR; amplification; probe; hybridisation;
XX detection; diagnosis; ds.

XX

XX Pseudomonas cepacia.

XX

XX JP07255486-A.

XX

XX 09-OCT-1995.

XX

XX 23-MAR-1994; 94JP-0051739.

XX

XX 23-MAR-1994; 94JP-0051739.

XX

XX (CANO) CANON KK.

XX

XX WPI; 1995-378541/49.

XX

XX Pseudomonas cepacia KK01 strain 16S rRNA gene - also related probes
XX and primers, useful for specific detection of P.cepacia strain KK01

XX

XX Claim 1; Page 21; 21pp; Japanese.

XX

XX This sequence represents the 16S rRNA gene of Pseudomonas cepacia
XX strain KK01. Fragments of the nucleic acid sequence (see T01872-T02316)
XX are useful as primers and probes for the specific detection of P.cepacia
XX strain KK01.

XX

XX Sequence 1526 BP; 382 A; 352 C; 486 G; 306 T; 0 other;

Query Match 54.5%; Score 805.4; DB 16; Length 1526;

Best Local Similarity. 75.3%; Pred. No. 1.2e-240;

Matches 1124; Conservative 0; Mismatches 326; Indels 43; Gaps 8;

Qy 2 atcatggtctagaactaaacgtgcggcgctgtttaaactgaaagtcgagcgggtagc 61

Db 9 atcctggctcagattgaacgtgcggcgatgctttacacatgcaagtcgaacggcagcac 68

Qy 62 aataactagc-----ggcgaacgggtgagtaaacacgtggtaattctctcc 107

Db 69 ggggtctgcacctggtggcgagtggtgcgaacgggtgagtaaacacatcggaactgtcctg 128

Qy 108 gsgtctgggataacttccgaaaggaaagctaataccggatagctcgttggatcacaaag 167

Db 129 tagtggggatagcccggaagcgggatttaaccgcatacgatcttcagat--gaaa 186

Qy 168 atttgatgggtaaaagtattctgttgagatgagcccgccgattagctagttggtg 227

Db 187 gcgggggaaccttcggcctcgctataggttggtgcgagtggtgattagctagttggtg 246

Qy 228 aggtaatggctcaccaagcgacgacgtgcgtaccggcgtgagaggtgtcccgccacaat 287

Db 247 gggtaaaagccctcaccaagcgacgacgtacgtagctggtgtgagagcagcagccacact 306

Qy 288 ggaactgagacacggttcccatctctacgggagggcagcagttlaagaactcttctcaatgg 347

Db 307 gggactgagacacgcccacagactcctacgggagggcagcagtggggaaattttggaacaatgg 366


```
QY 165 aagatttgatagtaagattttattgtctggagatgagcccgccggttagctagctagtg 224
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 189 agca--ggggaccttcggccttctgtagatgtagtgatgtagtgtagtgtagtg 246
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 225 gtgagataatggctaccacgaagcgacgacatcggttagcgcgcctcgagagggtgtccggccac 284
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 247 gtggggtaaaaggcctaccgaagcgacgacatccgttagctgtagtgtagtgtagtgtagtgtagtg 306
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 285 aatgggaactgagacacggtccatctctccacgagcgagcgagcaggttaagaatcttgcctca 344
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 307 accgggactgagacacgcccgaactcctacggagcgagcagtgagggaatcttgagacaa 366
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 345 tggggaaacctgaagcagcagccgcgcgtgaaacgaagaaggtcttcggattgaaagt 404
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 367 tggggcaacctgatccagcgcacgtgcgcgtgtgtgaagaagccctaggggtgtgaaagc 426
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 405 tcattaggcaggaa-----aaataagcagcaatgtgatgtgacgtgcct 451
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 427 acttcagtaggaggaaggcttcgggttaataccttgagtagctgtacgttacctacag 486
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 452 a--aagcaccgctactactgctccagcagccgcggttaactgattggtgcaagcgttgt 509
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 487 aagaagcaccggttaattctgtccagcagccgcggttaatacgaaggtgcgagcgttaa 546
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 510 tcggaatcattggcgtaaagggtgcgtaggcgcgatttgaagtcaggtgtgaaactgc 569
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 547 tcggaattactggcgtaaagcgcgtaggcgttttgaagtcagatgtgaaagcccc 606
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 570 gggctcaacccgtgcccctgacattgaaactacaagctcggagtttggagaggaagtg 629
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 607 gggctcaacccgtggaattgcatttgaactgcaagctagaatgcagtagaggagggtg 666
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 630 aattcagggtgagcgtgaaatgcgtgatatacttggaggaaacaccagtgcggaaggcga 689
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 667 aattccgggtgagcgtgaaatgcgtgatatacttggaggaaacaccagtgcggaaggc 726
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 690 ctgctggtctcaaaactgacgtgaggcagcaaaagcgtgtagtaaaacggagattagata 749
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 727 cctcctggactgacattgacgtgaggtgcgaagcgtgaggagcaaacaggattagata 786
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 750 ccccgtaataccagccctaaacgtgtctaccagttgttgggggtttta--accctcag 807
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 787 cccgtgtagtcacgcgttaaacagatgtctactagctgttgggaatcttagtattcttg 846
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 808 taacgaactaacggttaagttagaccgctgggactatctgcgaagatgaactca 867
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 847 tgacgaagttaacgcgataagttagaccgcttggggagtagcggcgcgaaggttaaaactca 906
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 868 aaggaattgacgggggtccgcacaaagcgttgagcatgtgtttaaattcgaatgatacccc 927
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 907 aatgaattgacggggccgcacaaagcgttgagcatgtgtttaaattcgaatgatacccc 966
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 928 aaaaacctcacctggctgacatgtagtca-atcatgtagagatatagaccttcgg 986
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 967 aagaaccttaccagcctgacatcccttggaaactttagagatagattggtgccttcgg 1026
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 987 --gacgattcacaggtgctgcatgtgtgtcgtcagctgctgtcgtgagattgtgggttaa 1044
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1027 gagccaagttagcaggtgctgcatgtgtgtcgtcagctgctgtcgtgagattgtgggttaa 1086
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1045 gtcccgcaacgagcgcaaacccctatcgtatgttgcta--ccttaagtgtggcactggatc 1102
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1087 gtcccgtaacgagcgcaaaccttgtccctagtgtgccgacacttcgggtgggaaactctagg 1146
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1103 gaaactgcggtgacaaacccggaagcggagcgtgacgtgacgtcaaatcctcatggtccttta 1162
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1147 gagactccggtgacaaacccggaagcgttggggagcagcgtcaatcctcatggtccttta 1206
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1163 tgtccaggccacacgctgtacaaatggccgacacagaggtgcgcaactcgaagagg 1222
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1207 cggcctgggtacacacgctgtacaaatgggcggttacagagggcagcgaagtcgcgaggcc 1266
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
QY 1223 aagctaattctaaagtcggtcccgagttcgagattgggggtctgcaactcgaccccatgaa 1282
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1267 gagcaaatcccttaaaacccgttcgtagtcggtgattggagtcgtcaactcgactccatgaa 1326
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1283 gtcgaaatcgctagtaaatcgcggtacagcagcgcgctgaatcagttcccggaaccttgt 1342
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1327 gtcgaaatcgctagtaaatcgcggtacagcagcgcgctgaatcagttcccggaaccttgt 1386
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1343 acacaccgcccgtcacacacactgagtgaggagcaccgcaagtggtcttgttaaccgta 1402
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1387 acacaccgcccgtcacacacactgagtgaggagcaccgcaagtggt-tagctaaaccttc 1445
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1403 aggaacacagactactaagtgaaactcgtaaaaggggtgaaagtcgtaacaagggtacc 1459
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1446 gggagacgataccacacggtgtggttcagtgagtggtggaagtcgtaacaagggtgagc 1502
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 15

```
T18765
ID T18765 standard; rRNA; 1536 BP.
AC T18765;
XX
XX T18765;
DT 05-JUL-1996 (first entry)
DE Pseudomonas testosteroni 16S ribosomal RNA.
KW Atrazine; pesticide degradation; soil decontamination;
KW bioremediation; s-triazine; herbicide; ss.
XX Pseudomonas testosteroni.
OS
XX
FH Location/Qualifiers
FT misc_difference 46
FT /*tag= a
FT /note= "base 46 is identified as 'n'"
XX
XX US5508193-A.
XX
XX 16-APR-1996.
XX
XX 31-AUG-1993; 93US-0114695.
XX
XX 31-AUG-1993; 93US-0114695.
XX
XX (MINU ) UNIV MINNESOTA.
XX
XX Mandelbaum RT, Wackett LP;
XX
XX WPI; 1996-208726/21.
XX
XX Biologically pure culture of atrazine-degrading Pseudomonas - useful
XX to detoxify atrazine, e.g. in soil, at a wide variety of concns.
XX
XX Example 2; Column 33-36; 34pp; English.
XX
XX Novel bacterial strain ADP (ATCC 55464), isolated from atrazine-
XX contaminated soil, is capable of degrading s-triazine cpds.,
XX including atrazine. In an attempt to identify the strain, the 16S
XX ribosomal RNA sequence (T18760) was compared to that of Escherichia
XX coli (T18759), Pseudomonas citronellolis ATCC 13674 (T18761-63),
XX Pseudomonas aeruginosa (T18764), Pseudomonas testosteroni (T18765)
XX and Pseudomonas cepacia (T18766). It was concluded that ADP is
XX closely related to, but distinct from, P. citronellolis and P.
XX aeruginosa.
XX
XX Sequence 1536 BP; 384 A; 355 C; 482 G; 314 U; 1 other;
SQ
```

Query Match 54.4%; Score 803.8; DB 17; Length 1536;
Best Local Similarity 60.2%; Pred. No. 3.6e-240;
Matches 897; Conservative 216; Mismatches 343; Indels 34; Gaps 7;

QY 1 gacatggctcagaactaacgctgcggc-gcgtattaaacatgcaagtcgagcgg---- 55
Db 15 gaucucgucagauagacgucgucgcaungcuuuaacacagcaagcuaacggaac 74
QY 56 -----ggtagcaatcacctagcggcgaaagcgggtgagtaacacgtggttaattctctccga 109
Db 75 agguucucggaugcugacgagcggcgaaacggyguguaaacaucggaacgucgucua 134
QY 110 gtcgtggataacttcccgaaggaaagtaataaccggatagtcctgttggatcacaaat 169
Db 135 gugggggaaauaacuacucgaaagagugcuauaacgcgaugagauacuacggaagaagca- 193
QY 170 ttgatagtaagaattattctgttgagatgagcccgccgagattagctagttggtgag 229
Db 194 -gggaccuucggccuugucuaagagcggcugugcagauuagguuuguggg 252
QY 230 gtaattggtcaccaaggcgacgacgtgtagccgctgagaggtgtccggccacaaagg 289
Db 253 guaaaggcuuaccagcugcgauucugugucgagagagacgacagccacacug 312
QY 290 aactgagacaggttccatctactcctcagggagcgacgagttaagaatcttgcataagg 349
Db 313 gacugagacagcccgacucucagcgaggcgagcaguggggaaauuuggacaaugggc 372
QY 350 gaaacccctgaagcagcgcgcgtgaacaaagaggtcttcgattgttaagttcatt 409
Db 373 gaaagcugacagccagcugcgcgagcaggaaggccuccugggguuuaaacugcuu 432
QY 410 aggcaggaaaaataagcag-----caatgtgatggtgacctgccta--aa 454
Db 433 uguacggaacgaaagccugggcuauauccccgggucaugacggguacgguagaa 492
QY 455 gcaccggctaaactagctgccagcagccgcgttaatacgtatggtcgaagcgtgttcgga 514
Db 493 gcaccggcuuacucagucagcagccgcgguuauacugagugcgaaagcguaauacugga 552
QY 515 atcattggggctaaagggtgcgtagcggatcttgaagtcaggtgtaaaactgcggct 574
Db 553 auuacgggcuuagagcugcgcgagcgguuuguaagacagugguuuaauccccgggcu 612
QY 575 caaccctggcctgcaactgaaactacaagtcgtgaggttgggagagcgaagtgaattc 634
Db 613 caaccugggaaacugccaugugacugcgaaggcuagugcgagcgagggggaugaauc 672
QY 635 cagggtgtagcgggtgaatgcgtagatctctggaggaacaccagtcggcgaaggcactgc 694
Db 673 cgcugugagcagugaaugcugauauugcgaggaacacccagugcgaaagcgaaucucc 732
QY 695 tggctcaaaactgacgtgaggcagcgaagcgtgggtagtaaacgggattagataccgcg 754
Db 733 ugggccugcacugacgcucaugcagaaagcuggggagcaaacaggauuagauaccucg 792
QY 755 gtaatccacgcctaaacgttgtaccagttgttgggggttttaacctcagtaacgaa 814
Db 793 guaguccacgcgccaauaagcuagucacugguuugguuuaacugacucaagaaacgaa 852
QY 815 cctaaccgattaaagtacccgctgggactatgctcgaagagtgaaactcaagggaat 874
Db 853 gcuaaacgcgugaagugaccccgugggagucgcccgaaggcuuagaaacuaaggaau 912
QY 875 tgacggggtccgcacaagcgtgagcatgtgttatttcgtatgataccccaacaaacc 934
Db 913 ugacggggaccgcacagcgggugaugugguuuaauucugcagcgcgaacaaacc 972
QY 935 tcacctgggcttgacatggaatcatgtagagatatatgaccttcgggcaga--- 991
Db 973 uuaccacuuugacauggcaggaacuuaccagagauuguuugugcugcaagaagaaacc 1032
QY 992 ---ttcacagggtgcatagggttgcgtacgctcgtcgtagatggtgggttaagtccc 1049
Db 1033 ugacacagggugcugcaguggcugucagcagcugcugugagugauguugguuuaagucc 1092
QY 1050 gcaacgagcgcaaacccctatcgtatgttgcctacctaagtgtgggcactggttcgaaactg 1109

Db 1093 gcaacgagcgcacaaccuugccaauuguguaucuuagugagcacucaauagggacug 1152
QY 1110 ccggtgacaaaaccggaggaagcggggtgacgtcaaatcctcatggcctttatgtccag 1169
Db 1153 ccggugacaaaaccggaggaaguggggaugacgucacgaugccuuaugguug 1212
QY 1170 gcccacacagctgctactaatggccgatacagagggtgcgcaactcgaagggagctaa 1229
Db 1213 ggcuaacacacgcuacuaauggcuguaacaaaggguugccaaaccgcgagggggagcuua 1272
QY 1230 tctctaaagtcggtcccagttccagttcgggtctgcaactcgaccccatgaagtcgaa 1289
Db 1273 ucccauaagccagucgugacgucgagcugcagucgacucgugugagucgaa 1332
QY 1290 tgcgtagtaatacgcggatcagcatccgcggtgaatacgttccccgaccttgcacacc 1349
Db 1333 ucgcuaaauucguggaucagaaucaagcgguaaauacguuucccgguucuaacacc 1392
QY 1350 gccctcacacacacccctgagtgagggggaccccgaaagtgtcttgttaaccgtaaggagac 1409
Db 1393 gcccgucacacacaugggagcgguucucgcgaagagug-guagccuaacgguaggagg 1451
QY 1410 agactactaagtgaaactcgttaaagggggtgaagtcgataacaaggtacc 1459
Db 1452 cgcuaaccacgcgggguucgugagcgggugagugggugaaugcuuaacaagguagc 1501

Search completed: March 27, 2001, 08:22:56

Job time: 5298 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March. 27, 2001, 06:53:33 ; Search time 132.3 Seconds
(without alignments)
1799.197 Million cell updates/sec

Title: US-09-380-826A-1

Perfect score: 1477

Sequence: 1 gatcatggctcagaactaac.....ccgtaaatgattcctgcag 1477

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 280836 seqs, 80580151 residues

Total number of hits satisfying chosen parameters: 561672

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	814.2	55.1	1536	US-08-642-229A-1	Sequence 1, Appli
2	805.4	54.5	1521	US-08-501-126-20	Sequence 20, Appl
3	803.8	54.4	1536	US-08-114-695A-7	Sequence 7, Appli
4	801.8	54.3	1512	US-08-995-960-2	Sequence 2, Appli
5	801.6	54.3	1555	US-08-737-653-160	Sequence 160, App
6	792.4	53.6	1517	US-09-248-528-2	Sequence 2, Appli
7	789.8	53.5	1516	US-09-248-528-3	Sequence 3, Appli
8	780.2	52.8	1542	US-08-114-695A-1	Sequence 1, Appli
9	780	52.8	1542	US-08-757-653-158	Sequence 158, App
10	779.8	52.8	1474	US-08-114-695A-8	Sequence 8, Appli
11	778.4	52.7	1556	US-08-995-960-1	Sequence 1, Appli
12	772.6	52.3	1484	US-08-632-470-53	Sequence 53, Appl
13	770.6	52.2	1518	US-08-114-695A-6	Sequence 6, Appli
14	768.4	52.0	1485	US-08-299-810A-27	Sequence 27, Appl
15	766.6	51.9	1542	US-08-875-445-21	Sequence 21, Appl
16	754.6	51.1	1494	US-08-632-470-49	Sequence 49, Appl
17	754.2	51.1	1452	US-08-642-229A-2	Sequence 2, Appli
18	753.6	51.0	1455	US-08-642-229A-3	Sequence 3, Appli
19	749.6	50.8	1464	US-08-938-858-1	Sequence 1, Appli
20	749	50.7	1508	US-08-632-470-44	Sequence 44, Appl
21	747.8	50.6	1452	US-08-276-943-1	Sequence 1, Appli
22	747.8	50.6	1452	US-08-716-841-1	Sequence 1, Appli
23	744.6	50.4	1430	US-08-902-518A-1	Sequence 1, Appli
24	743.8	50.4	1415	US-08-632-470-52	Sequence 52, Appl
25	741.8	50.2	1440	US-08-632-470-26	Sequence 26, Appl
26	741	50.2	1462	US-09-191-099-3	Sequence 3, Appli
27	740.2	50.1	1440	US-08-632-470-33	Sequence 33, Appl
28	740.2	50.1	1440	US-08-632-470-35	Sequence 35, Appl

29 740.2 50.1 1440 2 US-08-632-470-41 Sequence 41, Appli
30 739.8 50.1 1438 2 US-08-632-470-24 Sequence 24, Appli
31 738.6 50.0 1440 2 US-08-632-470-29 Sequence 29, Appli
32 738.4 50.0 1458 2 US-08-632-470-45 Sequence 45, Appli
33 737.8 50.0 1436 2 US-08-632-470-34 Sequence 34, Appli
34 735.6 49.8 1569 2 US-08-632-470-47 Sequence 47, Appli
35 735.4 49.8 1440 2 US-08-632-470-23 Sequence 23, Appli
36 735.4 49.8 1440 2 US-08-632-470-36 Sequence 36, Appli
37 735.4 49.8 1440 2 US-08-632-470-37 Sequence 37, Appli
38 735.4 49.8 1440 2 US-08-632-470-46 Sequence 46, Appli
39 735 49.8 1442 2 US-08-632-470-30 Sequence 30, Appli
40 734 49.7 1439 2 US-08-632-470-31 Sequence 31, Appli
41 733.8 49.7 1440 2 US-08-632-470-28 Sequence 28, Appli
42 733.8 49.7 1440 2 US-08-632-470-43 Sequence 43, Appli
43 731.6 49.5 1446 3 US-09-191-099-8 Sequence 8, Appli
44 731.2 49.5 1432 2 US-08-632-470-25 Sequence 25, Appli
45 731.2 49.5 1441 2 US-08-632-470-38 Sequence 38, Appli

ALIGNMENTS

RESULT 1
US-08-642-229A-1
; Sequence 1, Application US/08642229A
; Patent No. 5874291
; GENERAL INFORMATION:
; APPLICANT: Herwig, Russell P.
; APPLICANT: Bielefeldt, Angela R.
; APPLICANT: Stensel, H. David
; APPLICANT: Strand, Stuart E.
; TITLE OF INVENTION: Degradation of Environmental Toxins by a
; TITLE OF INVENTION: Filamentous Bacterium
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen O'Connor Johnson and Kindness PLLC
; STREET: 1420 Fifth Avenue, Suite 2800
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: WA 98101-2347
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/642,229A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/246,865
; FILING DATE: 20-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Sheiness, Diana K.
; REGISTRATION NUMBER: 35,356
; REFERENCE/DOCKET NUMBER: UOFW19233
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 682 8100
; TELEFAX: (206) 224 0779
; TELEX: 4938023
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1536 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: nucleic acid
; DESCRIPTION: "16S ribosomal DNA"
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Comamonas testosteroni ATCC No. 5874291 11996

TELEPHONE: (202) 887-1500
TELEFAX: (202) 887-0763
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 1521 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Bacillus sp.
INDIVIDUAL ISOLATE: DSM 8721
US-08-501-126-20

Query Match 54.5%; Score 805.4; DB 3; Length 1521;
Best Local Similarity 76.9%; Pred. No. 7.3e-281;
Matches 1097; Conservative 0; Mismatches 292; Indels 37; Gaps 8;

QY 69 agcgccgacggtgagtaaacagt-ggtaattcttccctccagctggtggataaactttccg 127
DB 84 AGCGCGGACGGGTGAGTAACACAGTGGCGCAACCTACCTTGTAGACTGGGATAAATCCCGG 143

QY 128 aaagaaagctaataaccggatagtcctgttggatcacagaatttgataggtaaaagatta 187
DB 144 AAACCGGGGCTAATACCGGATGATCAT-TTGGATCGCATGATCCGAATGTAAGCTGGGG 202

QY 188 ttg-----cttgagatgaagccgagcgccgcttagcttagctggtgaggttaattggc 237
DB 203 ATTATCTCAGTGCAGATGGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 262

QY 238 tcaccaagcgacgacgtgtagcgccgctgagaggtgtccgggccacaaatgaaactgaga 297
DB 263 TTACCAAGCGGACGATCGGTAGCCGACCTGAGAGGCTGATCGGCCACACTGGAACCTGAGA 322

QY 298 caagtcataactctacggagagcagcaggttaaaatcttgcataatgggggaaacct 357
DB 323 CACGCTCCAGACTCTACGGGAGGAGGAGGAGTGGGAATCTCCGCAATGGGGGAAAGCCT 382

QY 358 gaagcagcagcgcgctgaacgaagaggtcttcggttgaagttcattagcagga 417
DB 383 GACGCTGCAACCGCGGTGAAGATGAAGTTTTCGGATCGTAAGTTCTGTATGAGGG 442

QY 418 aaataagcagcaatgtgatgttaactgcct-----aaagcaccgg 461
DB 443 AAGAACAAAGTGCCTTGAATAGGTCCGACCTTGACGGTACCTCACGAGAAAGCCCGG 502

QY 462 ctactacgtgccagcagcgcggttaatactgattgtgcaagcgttgttcggaatcattg 521
DB 503 CTAACTAGTGGCCAGCGCGGTAATACGTAGGGGCAAGCGTTGTCCGGAATATTATG 562

QY 522 ggcgtaaaggtgctagcgagatttgaagtcaggtgtgaactgcgggtcaccgcg 581
DB 563 GCGGTAAAGCGCGCAGCGGCTCTTAAGTCTGATGTGAAGCCACCGGCTCAACCGT 622

QY 582 tggcctgcacttgaactacaagtctgagtttggagaggcaagtgaatccaggtgt 641
DB 623 CGAGGCTCATTTGAACTGGGGGACTTGAGTGTAGGAGGAAAGTGGAAATTTCCACGCT 682

QY 642 agcgtgaaatcgttagatctgagaaacaccagtgagcagcagcacttgcgtccta 701
DB 683 ACGGGTGAATACGTAGATATGTGGAGGAACACCACTGTCGGAAGCGACTTTCTGGCCTA 742

QY 702 aaactgagcgtgagcagcaagcgtgggtagtaaacgggattagatacccggttaatcc 761
DB 743 CAACTGACGCTGAGCGGCAAGCGTGGGGAGCAACAGAGATTAGATACCTGGTAGTCC 802

QY 762 acgcctaaacttbtctaccagttgttgggggttt--taacctgaagtaacgaacctaa 819
DB 803 ACCCGCTAAACGATGAGTCTAGTGTAGGGGTTTCGATACCTTGTAGTCCGCAAGTTAA 862

QY 820 cggattaaagtagacgccttggggaactatgctgcgaagagtgaactcaaaaggaattgacg 879

DB 863 CACATTAAAGCACTCCGCTGGGAGTACGGCGCAAGCTGAAGCTCAAGGAATTGACG 922
QY 880 ggggtccgcacacgagcgtgagcagcgtgtttaaattgatatacccaaaaacactacc 939
DB 923 GGGGCCCGCACAGAGTGGAGCATGTGTTTAATTCGAAGCAACGCGAAGAACCTTACC 982

QY 940 tgggcttgaca-----tggatctgaatcatgtagagatatatgagccttcgggcagattc 994
DB 983 AGGTCTTGACATCTCTGACACCTCTGGAGACAGAGCGTTCCCTTCGGGGGACAGAGTG 1042

QY 995 acagtgctgcatggttgcctcagctcgtgctgagatgttgggttaagtcccgcaac 1054
DB 1043 ACAGTGTGTCATGTTGTCGTCAGCTGTCGTGAGATGTTGGTTAAGTCCCGCAAC 1102

QY 1055 gagcgaacccct-atcgctattgttactcttaagtgtggcaactggtacgaaactgcgg 1113
DB 1103 GAGCGCAACCTTGATCTTAGTTGCCAGCATTCAGTTGGCACTCTAAGGTGACTGCCGG 1162

QY 1114 tgacaaacggagagcgggagtgatgacgtcaaatctctcatgacctttatgtccagggcc 1173
DB 1163 TGATAAACCGGAGGAGGTGGGATGACGTCAAATCATCATGCCCTTATGACCTGGGCT 1222

QY 1174 acacagtgctacatgcccgtacagagaggtgcgcaactcgcaagagggagactaatctc 1233
DB 1223 ACACAGCTGCTACAATGGATGGTACAAAGGCGAGACCGGAGGTTAAGCGAATCCC 1282

QY 1234 taaagtcggtcccgatctcggttgggtctgcaactcgaccccatgaagtgcggaatgc 1293
DB 1283 ATAAGCCATTCTCAGTTCGAGTTCAGGCTGCACTGCCTGCATGAAGCGGAATTGC 1342

QY 1294 tagtaatcgcgatcagcagcgcggtgaaatcacgttcccggacaccttgacacacggcc 1353
DB 1343 TAGTAATCGCGATCAGCATGCCGGGTGAATACGTTCCCGGGTCTTGTACACACGGCC 1402

QY 1354 gtcacacacactgagtgaggagacacacaaaggtcttctttaaaccgtaagagacagac 1413
DB 1403 GTCACACACAGAGAGTTGTAACACCGGAGTCG-GTGGGTAACTTTTGGGCGCAGCC 1461

QY 1414 tactaagtgaaactcgtaaaagggtgaagtcgtaacaaaggtacc 1459
DB 1462 GNCGAAGTGGGACAGATGATTGGGTGAAGTCGTAACAAAGGTATC 1507

RESULT 3
US-08-114-695A-7
Sequence 7, Application US/08114695A
Patent No. 5508193
GENERAL INFORMATION:
APPLICANT: Mandelbaum, Raphael T.
APPLICANT: Wackett, Lawrence P.
TITLE OF INVENTION: DEGRADATION OF S-TRIAZINES IN SOIL AND
TITLE OF INVENTION: WATER
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCHWEMAN, LUNDBERG & WOESSNER, P.A.
STREET: 3500 IDS CENTER
CITY: MINNEAPOLIS
STATE: MN
COUNTRY: USA
ZIP: 55402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/114,695A
FILING DATE: 31-AUG-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MUETING, ANN M.
REGISTRATION NUMBER: 33,977

```

; REFERENCE/DOCKET NUMBER: 600.268US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-339-0331
; TELEFAX: 612-339-3061
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1536 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: rRNA
; ORIGINAL SOURCE:
; ORGANISM: Pseudomonas testosteroni
; US-08-114-695A-7

```

```

Query Match      54.4% Score 803.8; DB 1; Length 1536;
Best Local Similarity 60.2% Pred. No. 2.8e-380;
Matches 897; Conservative 216; Mismatches 343; Indels 34; Gaps 7;

QY 1 gatcatggtcgaactaactgctggcgc-gcgtcttaacatcaagtcgagcgg---- 55
DB 15 GAUCCUGGUCAGAUAGAACGCGGCGCAUNGCUUACACAUAGCAAGUCGAACGUUAC 74

QY 56 -----ggtagcaataactagcggcgacgggtgagtaacacgtgtgtaattcttcga 109
DB 75 AGGUCUUCGGAGUCGACGAGCGCAACGGGUGAGUAUACACUGCAACGUCGUAGUA 134

QY 110 gtcctgggataactctccgaaggaaagctaataccgtagctgtgtgatacaagat 169
DB 135 GUGGGGAUAUAUCUGCAAGAGAGUACUAUACCGCAUGAGAUUCUACGGAUAAGCA- 193

QY 170 ttgatagtaaaagatttattgtctggtagagcccgccggtgattagttggtgag 229
DB 194 -GGGACCUUCGGCCUUGUCUACUAGAGCGGUGAGUGGAGAUUAGGUUGUGGG 252

QY 230 gtaatgctcacaagcgacgacatcggttagccgctgagagggtgtccggccacaatgg 289
DB 253 GUAAAGGCUUACCAAGGCGUGGUCUGUAGUGGUGGUGGACGAGGACGACACUGG 312

QY 290 aactgagacacggtccatactctacggggagcgagcttaagaattctgtcgaatggg 349
DB 313 GACUGAGACAGCGGCCAGACUCCUACGGGAGGACGAGUGGGGAUUUUGGACAAUAGGGC 372

QY 350 gaaacctgaagcagcagcgcgctgaacgaagaaggttcttgattgtaagttcatt 409
DB 373 GAAAGCCUGAUCCAGCAAUCCGCGUGCAGGAGUAGAGGCGCCUCUGGUGUUAACUGCUU 432

QY 410 aggcaggaaaaataagcag-----caatgtgatgtgtacctgccta--aa 454
DB 433 UGUACGGAACGAAAGCCUGGGGCUAAUUAUCCCGGUGUACUGACGGUACCGUAAAGAUAA 492

QY 455 gcaccggttaactacgttgcagcagccgcggttaatacgtatggtgcagcgtttgtcgga 514
DB 493 GCACCGCUAAUACUGGCGCAGCAGCGCGGUAUAGUAGUGGUGGCAAGCUUAUUCGGA 552

QY 515 atcattggcgtaaaaggtcgtaggcggtattgttaagtcagggtgtgaaacactcgggct 574
DB 553 AUUACUGGGCGUAAAGGUGCGGCGGCUUUUUGUAAAGACAGUGGUGUAAUCCCGGGCU 612

QY 575 caaccgtgctcctcacttgaatacctaagcttggagtttggagagcgcaagtggaaattc 634
DB 613 CAACUUGGGAACUGCCAUUGUAGUCGACGACGCUAGAGUGCGGACGAGGUGGAUUAUUC 672

QY 635 caggtgtagcgtgtaaatcgctagatatcttgagggaacacacagtcggcgaaggcgacttc 694
DB 673 CGGUGUAGAGUGAUAUAGUAGUAGUGGAGGAACACCGAUGGCGAAGGCAAUUCCCC 732

QY 695 tggctcaaaactgaactgagcagcaaaagcgtgtggttagtaaacgggattagataccccc 754
DB 733 UGGGCCUGCACACGUCUACGACGAAGGCGGUGGAGGACCAAGGAUUAUAGUACCCUG 792

QY 755 gtaatccacgcctaaactgttctaccagtgtgtggtgggggttttaaccctcagtaacgaa 814

```

RESULT 4

```

US-08-995-960-2
; Sequence 2, Application US/08995960
; Patent No. 6030818
; GENERAL INFORMATION:
; APPLICANT: PAGI, Michel
; APPLICANT: LANDRY, Nathalie
; APPLICANT: BOISSINOT, Maurice
; APPLICANT: HILIE, Marie-Claude
; APPLICANT: HARVEY, Mario
; APPLICANT: GAGNI, Martin
; TITLE OF INVENTION: BACTERIAL MASS PRODUCTION
; NUMBER OF INVENTION: OF TAXANES AND PACLITAXEL
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pike, Bronstein, Roberts & Cushman, LLP
; STREET: 130 Water Street
; CITY: Boston
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows

```



```

; SOFTWARE: FastSEQ for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/995,960
; FILING DATE: 22-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: RESNICK, David S
; REGISTRATION NUMBER: 34,235
; REFERENCE/DOCKET NUMBER: 1826/47986
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-523-3400
; TELEFAX: 617-523-6440
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1512 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: rRNA
; US-08-995-960-2

```

```

Query Match      54.3%; Score 801.8; DB 3; Length 1512;
Best Local Similarity 75.3%; Pred. No. 1.4e-279;
Matches 1134; Conservative 0; Mismatches 322; Indels 49; Gaps 9;

```

```

QY 1 gatacggctcagaactaacgctggcgcgcttcataaactgaagtcgaagcgggtag 60
DB 8 GATCATGGCTCAGGATGAACGCTGGCGGGTCCCTAAATACATGCAAGTCGACGGAATGGA 67
QY 61 caat-----acctagcgcggaacgggtgagtaacacgt-ggtaactcttc 103
DB 68 TTAAGAGCTTGCTCTATGAAGTTAGCGCGCAGCGGGTGAGTAACACGTGGGTAACTTAC 127
QY 104 ctccgagctcgggataactttccgaagaaagtaataccggatagtcctgttgatca 163
DB 128 CCATAAGACTGGGTAACCTCCGGGAAACCGGGCTTAATACCGGATAAATTTTGAACCTGC 187
QY 164 caagattgatagtaag-----attattgcttgagatgagcccgccgattca 216
DB 188 ATAGTTGCAAAATGAAGGCGGCTTCGGCGCTGCACCTATGATGATGACCCGCGTCGCATTA 247
QY 217 gctagttgtaggtaattggtccacaaagcgacgatcggttagccggcctgagaggggtg 276
DB 248 GCTAGTTGTTAGGTAACGGCTCACCAGCGCAGCATGGCTAGCCGACCTGAGAGGGTGA 307
QY 277 ccggccacaatggaactgagacacggttccatactcctcagggagcagtgtaagaatc 336
DB 308 TCGGCCACACTGGGACTGAGACACGCCGCCAGACTCCTACGGGAGGCGAGCAGTAGGGAATC 367
QY 337 ttgcctcaatggggaaacccctaaacagcagcgcgcgctgaacgaagaaggtcttcgat 396
DB 368 TTCCGCAATGGACGAAAGTCTCAGCGAGGCAACCGCGCTGAGTGAGAAAGGCTTTCGGGT 427
QY 397 tgtaagttcattagcaggaataaagcagcaatgtgatgtagtacctgcct----- 451
DB 428 CGTAAACTCTGTTTAGGGAGACACAGTCTGCTAGTTGAATGAAGCTGGCACCCTTGACGG 487
QY 452 -----aaagcaccggtaactacgtgccagcag-cgcggtgaatacgtatggtg 499
DB 488 TACCTAAACAGAAAGACACGGCTAACTACCTGCCAGCAGCCCGGTAAATACGTAGGTGG 547
QY 500 caagcgttctcgaatcatattggcgtaaaaggtgcgtaggcggatttgtaagtcaggtg 559
DB 548 CAAGCGTTATCCGAATTTATGGCGGTAAAGCGCGCGCAGGTGTTCTTAAGTCTGATG 607
QY 560 tgaactcgggctcaaccctggcctgacctgaactacaagctcgtgatttgggag 619
DB 608 TGAAGCCCGCAGGCTCAACCGTGGAGGGTCAATGGNAACCTGGGAGACTTGATGGCAAG 667

```

```

QY 620 aggaagtggaattccaggtgtagcggtagaactgagtagatctatctggaggaacaccagtg 679
DB 668 AGGAAAGTGGAAATTCATGTGTAGCGGTCAAATGCGTACAGATATATGAGGAACACACCAGTG 727
QY 680 gcaagggcgacttgctgctcaaaaactgacgctgagggcagcaagcgtgggtagtaaacg 739
DB 728 GCGAAGGCGACTTTCTGGTCTTAACACTGACACTGAGGCGCGAAAGCGTGGGAGCAACA 787
QY 740 ggattagatcccccggttaatccacgctcaaacgttgctaccagttgttgggggtttc- 798
DB 788 GGATTAGATACCTGGTAGTCCACGCCGTAAACGATGAGTGTCTAAGTGTAGAGGGTTTC 847
QY 799 -aacctcagtaaacgaacctaacggattaaagtagacgcctggggactatgctcgcaaga 857
DB 848 CGCCCTTTAGTGTCTGAAGTTAAGCATTAAGCACTCCGCTGGGAGTAGCGGCCGCAAG 907
QY 858 gfgaaactcaaggaattgacgggggtccgcacaaagcgtgagcagtggttttaattcg 917
DB 908 CTGAAACTCAAGGAATTAAGCGGGGCCCGCACAGCGGTGGAGCATGTGGTTTAATTCG 967
QY 918 atgatacccccaaaacctcacctgggcttgacatggatcctgaatcatgtagagatatag 977
DB 968 AAGCAACGCGAAGAACCTTTACCAGGCTCTTGACAT-CCTCTGAAAACTCTAGAGATAGAGC 1026
QY 978 agc-----cttcggggcagattccacaggtgctgcatggttgcagctcgtgctgaga 1033
DB 1027 TTCTCCTCGGGAGCAGAGTAGCAGGTGTGATGTTGCTGCTCAGCTCCTGCTGCTGAGA 1086
QY 1034 ttttgggttaagttccgcgaacgagcgaacccct-atcgtatgttgcaccttaattgg 1092
DB 1087 TGTGGGTTTAAAGTCCCGCAACGAGCGCAACCCCTTGATCTTAGTTGCCATCATTAAGTTGG 1146
QY 1093 gcactggttacgaaactgccggtgacaaacggaggaagcggggagtgacgtcaaatccctc 1152
DB 1147 GCACCTTAAGGTGACTGCCGCTGACAAACCGGAGGAAGTGGGGATGACGTCAATCATC 1206
QY 1153 atggcctttatctcagggccacacagctgctacaaatggccgatacagagggctgcgaac 1212
DB 1207 ATGCCCTTTATGACCTGGGCTACACACGTGCCACAAATGGACGCTACAAAGAGCTGCAAGA 1266
QY 1213 tcgaagagggagcctaactctctaaaagtcggtcccgactggaattgggtctgcaactcg 1272
DB 1267 CCGCAGGTGGAGCTAATCTCAATAAACCCCTTCTCAGTTGCGGATTTAGCTGCAACTCG 1326
QY 1273 acccatgaagtcggaatcgtagtaacgcgcatgacgcggtgacgcggtgaaacgtctcc 1332
DB 1327 CCTACATGAAGCTGGAATCGTAGTAATCGCGGATCAGCATGCCGCGGTGAATACGTTC 1386
QY 1333 cggacctgtcacaccgcccgttcacacacactgagtgggggagcaccggaagtggcttt 1392
DB 1387 CGGGCCTTTACACACGCCCGCTCACACACGAGAGGTTTGTAAACCCCGAAGTCGGTGGG 1446
QY 1393 gttaacctgaaggaacagactactaaagtgaacctcgtaaaaggggtgaaagtcgtaaca 1452
DB 1447 GTAACCTTTATGGAGCCCGCCCTAAGGTGGGACAGATGATTTGGGGTGAAGTCGTAACA 1506
QY 1453 agta 1457
DB 1507 AGGA 1511

```

```

RESULT 5
US-08-757-653-160
; Sequence 160, Application US/08757653
; Patent No. 5843669
; GENERAL INFORMATION:
; APPLICANT: Kaiser, Michael W.
; APPLICANT: Lyamichev, Victor I.
; APPLICANT: Lyamichev, Natasha
; TITLE OF INVENTION: Cleavage Of Nucleic Acid Using
; TITLE OF INVENTION: Thermostable FEN-1 Endonucleases
; NUMBER OF SEQUENCES: 190

```


GENERAL INFORMATION:

; APPLICANT: Oriol, Patrick J
 ; APPLICANT: Padmakumar, Rugmini
 ; APPLICANT: Kim, Sang H
 ; TITLE OF INVENTION: Method for Producing Amide Compounds Using a Nitrite
 ; FILE OF INVENTION: Hydratase from a Thermophilic Bacillus
 ; FILE REFERENCE: MSU 4.1-401
 ; CURRENT APPLICATION NUMBER: US/09/248,528C
 ; CURRENT FILING DATE: 1999-02-10
 ; EARLIER APPLICATION NUMBER: 60/083,485
 ; EARLIER FILING DATE: 1998-04-29
 ; NUMBER OF SEQ ID NOS: 18
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 2
 ; LENGTH: 1517
 ; TYPE: DNA
 ; ORGANISM: Bacillus sp.
 ; FEATURE:
 ; NAME/KEY: rRNA
 ; LOCATION: (1)..(1517)
 ; PUBLICATION INFORMATION:
 ; DATABASE ACCESSION NUMBER: 226929
 ; DATABASE ENTRY DATE: 1998-07-02
 ; US-09-248-528-2

Query Match 53.6%; Score 792.4; DB 3; Length 1517;
 Best Local Similarity 75.0%; Pred. No. 3.7e-276;
 Matches 1128; Conservative 0; Mismatches 326; Indels 50; Gaps 9;

QY 6 tggctcagacaaacgctgagcgcgctcttaaacatgcaagtcagcgggtagcaata 65
 DB 1 tggctcagacaaacgctgagcgcgctcttaaacatgcaagtcagcgggtagcaata 60
 QY 66 cc-----tagcgcgacgaggtgagtaaacgctggaatacttctctccg-ag 110
 DB 61 gcttgcctcttaggttagcgcgagcgggtgagtaaacgctggaatacttctctccg-ag 120
 QY 111 tctgggataacttccgaaagaaagctaataccggagtagctctgttgatcacaaagatt 170
 DB 121 actggataacttcggaaacggagctaataccggagtagctctgttgatcacaaagatt 180
 QY 171 tgataggtaag-----attatctgttgagatgagccgcgagcgttagctagt 222
 DB 181 tgggttgaaagcggcttttagctctcaactgagatggccgcgagcgttagctagt 240
 QY 223 tggtaggttaattggctcaccagggagcagctggttagcggcgtcgtgagaggtgtccggcc 282
 DB 241 tggtaggttaacggtcaccagggagcagctggttagcggcgtcgtgagaggtgtccggcc 300
 QY 283 acaatggaactgagaaacggtccatactctacggaggagcagcaggttaagaattctgtctc 342
 DB 301 acactgggactgagacacgcccagactctcagggagggcagcagtagggaattcttccgc 360
 QY 343 aatggggaaacccctgaagcagcagcgcgctgaaacgaaggtcttcgagattgtaaa 402
 DB 361 aatggacgaagtctgacggagcaacgcgcgctgagcgaaggtcttcgagattgtaaa 420
 QY 403 gttcattaggcaggaataaagcagcaattgta-----tgatgggtacctg 448
 DB 421 gctctgttgcaggaagaacagtagtacctctcgaacagggcggtaccttgacgggtacctg 480
 QY 449 ccta--aagcaccggctaaactcgtgccagcagcgcggttaatacgtatggtgcagcgt 506
 DB 481 acgaggaacccagcgttaactcgtgccagcagcgcggttaatacgtatggtgcagcgt 540
 QY 507 tggtaggaatcattgggcgttaaggttgcgtagggagatttgaagtcagggtgtgaaac 566
 DB 541 tggcgaagaattatgggcgttaagcgcgcgcgagggcgtcccttaagtcgtgattgaaac 600
 QY 567 tgcgggtcacaacccggcgtcactgcaactgaaactacaagctgaggttgggagagcgaag 626
 DB 601 tgcgggtcacaacccggcgtcactgcaactgaaactggaacttggagtagcagggagggag 660

QY 627 tggaaattccaggtgtagcgggtgaaatgcgtagatctctgaggaacaccagtcggcgaagg 686
 DB 661 cggaaattccaggtgtagcgggtgaaatgcgtagatctctgaggaacaccagtcggcgaagg 720
 QY 687 cgaactgtcgtgctcaaaactgacgtgacgacacgaacgctgggtagtaaaacgggattag 746
 DB 721 cggctctctggtcgtgtaactgacgctgagcgcgcaaaacgctgggagcgaacgattag 780
 QY 747 ataccgggtaataatccacgcccctaaacgctgtctaccagttgttgggggtttt--aacctt 804
 DB 781 atactctgtagtccacgcccgttaaacgtagtgcctaaaggttagaggtatccaccctt 840
 QY 805 cagtaacgaacctaagcagattaaagtagaccgctggggactatgctcgaagaagtgaaac 864
 DB 841 tagtctgcagcaaacgcattaaagcactccgctgggagtagcgcgcaagcgtgaac 900
 QY 865 tcaaaaggaattgacgggggtccgcacaaacgctgagcagctggttttaattcagatgac 924
 DB 901 tcaaaaggaattgacgggggtccgcacaaacgctgagcagctggttttaattcagatgac 960
 QY 925 cccaaacacctcacctgggcttgacat-----ggatctgaatcatgtagagatatag 977
 DB 961 gcgaagaaccttaccaggtcttgacatccctgacaaacctagagatagggcgttccctt 1020
 QY 978 agccttcgggcagattcacaggtgctgcatggttctgctcagctcgtgctgagatgatt 1037
 DB 1021 ttcgggggacaggggtgacaggtggtgcatggtgctgctgctgctgctgagatgatt 1080
 QY 1038 ggggttaagtcctcgcacacgacgcaacccctatcgtatgttctga-ctttaagttgggac 1096
 DB 1081 ggggttaagtcctcgcacacgacgcaacccctatcgtatgttctga-ctttaagttgggac 1140
 QY 1097 tggtaggaactgcccgggtgacaaacgggagggcggtgagcgtcaaatctctcatg 1156
 DB 1141 tctaaagtgactgcgggctaaaagtcggaggaaggtggggatgacgtcaaatcatcagc 1200
 QY 1157 cctttatgtccagggccacacacgctgctcaaatgcccgtacagagggctgcacaactcgc 1216
 DB 1201 cctctatgacctgggtctacacacgctgctcaaatgcccgtggtggtacaaagggcagcgaaccgc 1260
 QY 1217 aagagggaggaactatctttaaagtcgggtcccaggtcgggttgggtctgcactcgaacc 1276
 DB 1261 gagggtgagcgaatcccccaaaacacactctcagttcggattgacagctgcaactcgcctg 1320
 QY 1277 catgaagtcgggaactcgtagtaactcgggacagcagcgtgcagcgtgcagcttcccgga 1336
 DB 1321 catgaagtcgggaactcgtagtaactcgggacagcagcgtgcagcgtgcagcttcccgga 1380
 QY 1337 cctttgacacacgcccgttcacacacactgagtgaggagcaccggaagtgggtttgta 1396
 DB 1381 tcttgacacacgcccgttcacacacagaggtttgttaacacccggaagtgggtgta 1440
 QY 1397 accgta-aagagacagactactaaagtgaaactcgtaaaaggggtgaaagtcgtaacaaag 1455
 DB 1441 ccttcagggccagcccgccgaaggtgggcaaatgattgggtgaaagtcgtaacaaag 1500
 QY 1456 tacc 1459
 DB 1501 tagc 1504

RESULT 7

; US-09-248-528-3
 ; Sequence 3, Application US/09248528C
 ; Patent No. 6153415
 ; GENERAL INFORMATION:
 ; APPLICANT: Oriol, Patrick J
 ; APPLICANT: Padmakumar, Rugmini
 ; APPLICANT: Kim, Sang H
 ; TITLE OF INVENTION: Method for Producing Amide Compounds Using a Nitrite
 ; TITLE OF INVENTION: Hydratase from a Thermophilic Bacillus
 ; FILE REFERENCE: MSU 4.1-401

```

: CURRENT APPLICATION NUMBER: US/09/248,528C
:
: CURRENT FILING DATE: 1999-02-10
:
: EARLIER APPLICATION NUMBER: 60/083,485
:
: EARLIER FILING DATE: 1998-04-29
:
: NUMBER OF SEQ ID NOS: 18
:
: SOFTWARE: PatentIn Ver. 2.0
:
: SEQ ID NO 3
:
: LENGTH: 1516
:
: TYPE: DNA
:
: ORGANISM: Bacillus pallidus
:
: FEATURE:
:
: NAME/KEY: rRNA
:
: LOCATION: (1)..(1516)
:
: PUBLICATION INFORMATION:
:
: DATABASE ACCESSION NUMBER: Z26930/GenBank
:
: DATABASE ENTRY DATE: 1997-05-14
:
: US-09-248-528-3

```

Query Match	53.5%	Score	789.8	DB 3	Length	1516
Best Local Similarity	75.0%	Pred. No.	3.2e-275			
Matches 1124	Conservative	0	Mismatches	327	Indels	48
QY	9	ctcagaactaacctgctgcggcgctctttaaacaatgcaagtgcagcggtgtagcaatacct	68			
Db	1	ctcagacgaacgtgcggctgcctaatacatgcaatgcagtcgagcggaagggagct	60			
QY	69	ag-----cggcgacgggtgagtaaacgctggtaactcttctccg-agtct	113			
Db	61	tgtctcttaggttaacgcgcggcggtgagtaacacgtgggcaacctgccctgcagact	120			
QY	114	gggataacttcgaaagaaagactaataccggatgctgttgatcacagatttga	173			
Db	121	gggataacttcggaaaccggagctaataccggataaacccgaaaccgatgtttcg	180			
QY	174	taggtaaag-----attattgcttgagatgagccgcgcgcgtagctagtgtgg	225			
Db	181	gttgaaggcggttttagctgcactgcagatggcgccgcgcgcatagctagtgtgg	240			
QY	226	tgaggtaattggtccacaaaggcgacgatcgttagcgcgcctgagaggctgcgcgcaca	285			
Db	241	tgaggtaacggtccaaaggcgacgatcgttagcgcgcctgagaggctgcgcgcaca	300			
QY	286	atggaactgagacacggtccatactctacggagcgacgagttaagaattctgtcaat	345			
Db	301	ctgggactgagacagccagactctacggagcgacgtaggaaattcttcgcaat	360			
QY	346	gggggaaacctgaagcagcgcgcgtgaacgaagaaggtcttcggattgtaaaagt	405			
Db	361	ggcgaagcttgacgagcaacgcgcgtgagcgaagaaggtcttcggatcgtaagct	420			
QY	406	cattagcgaggaaaaatagcagcaatgga-----tgatggctacctgct	451			
Db	421	ctgtgttcaggaaagaacaagtgcgcttcgaacagggcggtaccttgacggctacctgacg	480			
QY	452	a--aegcacccgctaactacgtgccagacgcgcgcggttaatacgtatggtgcgaagcgttgt	509			
Db	481	aggaaegccagcctaactacgtgccagacgcgcggttaatacgtatggtgcgaagcgttgt	540			
QY	510	tcggaatacttgggcgttaaaagggtgcgttaggcgcgatttgttaagtcaggtgtgtgaaactgc	569			
Db	541	ccggaataattgggcgttaaaaggcgcgagcggtccttaagtcaggtgtgtgaaactgcg	600			
QY	570	gggctcaacccgtggcctgcaactgaaactacaagcttggagttgggaagagcgaagtgg	629			
Db	601	cggtcaacccgcgcgcgccttggaactgggaactgtgagtcagagaggggagcgg	660			
QY	630	aattccaggttagcgggtgaaatcgctagatatctggaggaaacacacagtgccgagagcgga	689			
Db	661	aattccacgttagcgggtgaaatcgctagagatgtggaggaaacacacagtgccgagagcg	720			
QY	690	cttgcgtctcaaaactgacgctgaggcacgaaagcgtgggtagtaaacgaggttagata	749			

Db	721	ctctctg	ccgttaactg	acgtgagcgcgaaagctggggagcgaacagattagata	780
Qy	750	cccggtaatccacgc	ccctaaacgttgtctaccagttgtgtgggggtttt--aaccttcag	807	
Db	781	ccctggtag	lccacgcgtaaacgataagtgctaagtttagaggggtatccaccctttag	840	
Qy	808	taacgaacctaa	cggattaaagttagaccgcctggggaactatgctgcgaagatgaaactca	867	
Db	841	tgtgtcagcaaac	gcgttaagcactccgctcgtgggagtacgcgcgaaggtgaaactca	900	
Qy	868	aaggaaattgac	gggggtccgcacaaagcgttgagcatgtgtttaattccatgataccc	927	
Db	901	aaggaaattgac	ggggccgcacaagcgttgagcatgtgtttaattccgaagcaagcg	960	
Qy	928	aaaaacctca	ctgggcttgacatggatctgaatc----atgtagagatatatgacct	982	
Db	961	aagaaccttacc	aggtcttgacatcccttgacaaccttagagatagggcgttccctttc	1020	
Qy	983	tcgggcagatt	ccaggtgtcgtcatgttgttcgtcagctcgtgtcgtaaatgttggtt	1042	
Db	1021	ggggacaaggt	gacagtggtcatgttgttcgtcagctcgtgtcgtagagatctggggtt	1080	
Qy	1043	aagtcgcgaac	agagcgcaaacccctatcgtatgtttgcta-ccttaagttgggcactggta	1101	
Db	1081	aagtcgcgaac	agcgcaaccttgacctagttgcagcatctcagttgggcactctaa	1140	
Qy	1102	cgaacctgc	cggtgacaaacggaggaagcgggatgacatcaaatccctcatggccttt	1161	
Db	1141	ggtagctgc	cgctaaaagtccggaggaaggtggggatgacgccaaatcaatgcacctt	1200	
Qy	1162	atgtccaggg	ccacacacgtgtcataatggccgatcacagaggggtcgccaactcgcgaag	1221	
Db	1201	atgacctgg	gtacacacgtgtcataatgggttggtacaaaagggcgaacgcgaggt	1260	
Qy	1222	ggagctaa	tctctaaaagtcggtcccgatcggatgtgggtctgcaactcgaccatga	1281	
Db	1261	ggagcaat	ccccaaaaaacactctcagttcggaatgcaggtgcgaactcgcgtcgatga	1320	
Qy	1282	agtcggaat	cgtagtaatcgcggtcagcatgcgcgggtgaaatcagttcccggaccttg	1341	
Db	1321	agcggaa	tgcgtagtaatcgcggtcagcatgcgcgggtgaaatcagttcccggaccttg	1380	
Qy	1342	tacacacc	cccggtcacaccactgagtggggaacacccaagtgctttgttaaccgt	1401	
Db	1381	tacacacc	cccggtcacaccacagagattgttaacaccgcgaagtcgggtggggtaacctt	1440	
Qy	1402	a-agggag	cacagactactaaggtgaaactcgttaaggggggtgaagtcgtacaagaagttacc	1459	
Db	1441	acggggag	cagccgcgaagtggggacaaaatgattgggtgaaatcgtacaagaagttacc	1499	

RESULT 8
US-08-114-695A-1
; Sequence 1, Application US/08114695A
; Patent No. 5508193
; GENERAL INFORMATION:
; APPLICANT: Mandelbaum, Raphael T.
; APPLICANT: Wackett, Lawrence P.
; TITLE OF INVENTION: DEGRADATION OF S-TRIAZINES IN SOIL AND
; TITLE OF INVENTION: WATER
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCHWEGMAN, LUNDBERG & WOESSNER, P.A.
; STREET: 3500 IDS CENTER
; CITY: MINNEAPOLIS
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

```

: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/114,695A
: FILING DATE: 31-AUG-1993
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: MUETING, ANN M.
: REGISTRATION NUMBER: 33,977
: REFERENCE/DOCKET NUMBER: 600,268US1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 612-339-0331
: TELEFAX: 612-339-4061
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1542 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: rRNA
: ORIGINAL SOURCE:
: ORGANISM: Escherichia coli
: US-08-114-695A-1

```

Query Match	52.8%	Score 780.2	DB 1	Length 1542
Best Local Similarity	60.0%	Pred. No. 9.6e-272		
Matches	900	Conservative 217	Mismatches 338	Indels 44
QY	1	gøtatggtcagaaactaaøctgøcgøcgøctttaaøcatcaøatcgaøtgøagøggøtag	60	
Db	15	GAUC AUGGUCUGAUAUGNACCGUGGCGGACGCCUACACAUUGCAANGUCGACGGUAACA	74	
QY	61	caøtac-----ctagøggøcgaøcggøtgatøaaøcøtgøtaøtø	101	
Db	75	CGGAAGAAGACUUGUCUUCUGACGAGUGGCGGACGGUGUAUAUGUCUGGGAAACU	134	
QY	102	tcøtcøggøctgøgøtaøacttøcøgaøaøgaøcøtaøtaøcøgøatøgøctøgtøgat	161	
Db	135	GCCUAGUUGAGGGGUAUAUCUAGUAAACGGUAACGUUAUAUACCGGAUA---AGCUGCGAG	191	
QY	162	cacaaøattøgatøggøtaøaøatøttøgtøtøgøatøgøcøcgøcøcøatøtagøtag	221	
Db	192	ACCAAGAGGGGGACCUUGGGGCCUUCUCCAUUGGCAUUGCCCCAGUUGGGAUAUAGCUAG	251	
QY	222	ttøgtøagøgtøaøtgøctøcøaaøagøcøgaøctøggøtagøcøggøctøgøaøggøtgøccøggøc	281	
Db	252	UAGUGGGGUAACGGGUCUACUAGCGGACGACUAUCCUACUGGUCUGACGAGGAUAGACCAGC	311	
QY	282	cacaaøtøgaøctøgøacøcøgtøcøatøcøtøcøggøggøcøgøcøagøcøagøtøtaøgaøatøgtøct	341	
Db	312	CACACUGNACUGAGACACGGUCCAGACUCCACGGGAGGACGACAGUGGGNAUAUUGCA	371	
QY	342	caaøtøggøggøaaøccøtøgaøgcøagøcøgcøcgøtgøaaøcøgaøaøggøtøctøggøatøgtøaa	401	
Db	372	CAAUUGGCGCAAGCCUGAUGCAGCGCAUCCGCGGUGUAUGAAGAGGCCUUCUGGGUUGUA	431	
QY	402	agøtøcøatøagøcøggøaaøaaøtaøaøcøgøcaøatøgtøa-----tgøatøggøtacøgt	448	
Db	432	AGUAUUUACCGGGGAGGAGGGAGUAUAAGUUUAUUACUUUUGCUUAUUGACUUUACCCG	491	
QY	449	cøcta--aøgcøacøggøcøtaøactøacøtgøcøagøcøgcøggøtgøaøcøtgøtgøcøaøggøt	506	
Db	492	CAGAAGAAGACACGGCUNACUCCGUCGACGACGCCGCGUAAUACGGAGGGUGCAAGCGU	551	
QY	507	tgøtøcøgaøaaøtøatøggøcgøtaøaaøggøtgøcgøtagøcøggøatøtgøtaøgtøcøagøtgøtaøaaøc	566	
Db	552	UUAUCGGAUUUAUCUGGGCGUAUAACGACGACGACGGUUGUUAUAGUCUAUGUUGAAUUC	611	
QY	567	tgøggøgøctøcaøccøcgøtgøggøctøgøactøgøaaøcøtaøcøaøgtøctøggøagøtgøggøagøcøaøg	626	
Db	612	CCGCGGCUACACUUGGNAUCUUAUCUUAUACUUGGCAAGCUUAUGACUUGACUUGAGCGGGG	671	
QY	627	tgøgaøttøccøagøgtøtagøggøtgøaaøatøcgøtagøatøctøtgøggøggøaaøcøcøagøtgøggøaaøgg	686	

Db	672	UAGAAUUCGAGUGUAGCGGUGAUAUUGCGUAGAGAUUGGAGGAUAUACCCGUGGCGGAAGG	731
Qy	687	cgactgtcgtctcaaaactgacgtctgagcgacgaagcgtgggtagtaaacgggattag	746
Db	732	CGGCCCCUGGAGCAAGACAGUCAGCUCAGGUCGGAACCGUGGGAGCAACACAGAUUAG	791
Qy	747	atccccggtaatccacgcgcctaaacgttgtctac-cagttgttgggggttttaaccctc	805
Db	792	AUACCCUGGUGAUUCCACGCGGUAACGAGUUGCGAGUUGGAGGUGGCGUUGAGCGGUG	851
Qy	806	agtaacgaactaacgattaaagttagacgcctggggactatgctgcgaagagtgaact	865
Db	852	GCUUCCGGAGCUAACGCGUUAAGUCGCGCCUGGGGAGUACGCCGCCAGGUUAAACU	911
Qy	866	caaaagaaatgcagggggtccgcacaacgggtggagcatgtggttttaactcgatgatacc	925
Db	912	CAAAAGAAUUGACCGGGGCCGCACAAAGCGUGGAGCAUGUGGUUUAAUUCGAUCCACAG	971
Qy	926	ccaaaaactcaactggctggcttgacatggaatcat-gtagagatatatagaccttc	984
Db	972	CGAAGAACCUUACCGUCUUGACAUCCACGCAAGAUUUCAGAGAUGAGAAUGUGCCUUC	1031
Qy	985	ggg--cagattcaacaggtgctgcattgtgtcgtcagctcgtgtcgtgagatgttggtt	1042
Db	1032	GGGAACCGUGAGACAGGUGCUGCAGGCGUCGUCGAGCGUCGUGGUGGAAUUGUGUGGU	1091
Qy	1043	aagtcctcgcaacgagcgcaaccctcatcgtatgtgtctacc--ttaagttgggcactggt	1100
Db	1092	AAGUCCCGCAACGAGCGCAACCCUUAUCCUUUGUUGCCAGCGGUCGCCGCGGAACUCAA	1151
Qy	1101	acgaactcgcggtgtgacaaacccggaggaaggcggggatgacgtcaaatctctcatggcctt	1160
Db	1152	AGGAGACUGCCAGUAAUACUGGAGGAGGUGGGAUGACGCUCAAGUACAUGGCGCCU	1211
Qy	1161	tatgtccaggccacacacagctgctcaaatgcccatacacagaggtgcgcaactcgcaaga	1220
Db	1212	UACGACACGGGCUACACACGUGCUACAUGGCGUAUACAAAGAAAGAACGACCGCGGAGA	1271
Qy	1221	ggagctaaactctaaagtcggtccagttcggtgattgggtctcgtcaactgcaccccatg	1280
Db	1272	GCAAGCGACCUCAUAAAGUCCGUCUAGUCCGGAUUGGAGUUCGACUCCACUCCUAUG	1331
Qy	1281	aagtcggaatcgttagtaatcgcggatcagcatcgcgcggtggaatacgtttcccggaactt	1340
Db	1332	AAGUCGAAUCGCUAGUAAUUGUGGAUCAAGAUGCCACGGUGAAUCGUUCCGGGCUU	1391
Qy	1341	gtacacaccgctgcacacaccactgagtggtggagcaccgcgaagtgtcttgttaaccg	1400
Db	1392	GUACACACCCCGGCUACACCAUGGAGUGGUGGUGUUGCAAAAGAAUGAG-GUAGUCUUAAACU	1450
Qy	1401	taeaggacagactactaagtgaaactcgtataaggggtgaagtcgttaacaggtacc	1459
Db	1451	UCGGAGGGCGCUUACCAUUUGUAUUCAUCAUGGCGGUGAAGUUGUUAACAGGUAAAC	1509
RESULT 9			
US-08-757-653-158			
; Sequence 158, Application US/08757653			
; Patent No. 5843669			
; GENERAL INFORMATION:			
; APPLICANT: Kaiser, Michael W.			
; APPLICANT: Lyamichev, Victor I.			
; APPLICANT: Lyamichev, Natasha			
; TITLE OF INVENTION: Cleavage Of Nucleic Acid Using			
; TITLE OF INVENTION: Thermostable FEN-1 Endonucleases			
; NUMBER OF SEQUENCES: 190			
; CORRESPONDENCE ADDRESS:			
; ADDRESSEE: Medlen & Carroll, LLP			
; STREET: 220 Montgomery Street, Suite 2200			
; CITY: San Francisco			
; STATE: California			
; COUNTRY: United States of America			
; ZIP: 94104			

DEBIT

US-08-757-653-158
; Sequence 158, Application US/08757653
; Patent No. 5843669
; GENERAL INFORMATION:
; APPLICANT: Kaiser, Michael W.
; APPLICANT: Lyamichev, Victor I.
; APPLICANT: Lyamichev, Natasha
; TITLE OF INVENTION: Cleavage Of Nucleic Acid Using
; TITLE OF INVENTION: Thermostable FEN-1 Endonucleases
; NUMBER OF SEQUENCES: 190
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States Of America
; ZIP: 94104

US-08-632-470-53
: Sequence 53, Application US/08632470
: Patent No. 5976791
: GENERAL INFORMATION:
: APPLICANT: MABILAT, CLAUDE
: APPLICANT: RAOULT, DIDIER
: TITLE OF INVENTION: NUCLEOTIDE FRAGMENT'S CAPABLE OF
: TITLE OF INVENTION: HYBRIDIZING SPECIFICALLY TO RICKETTSIA RDNA OR RRNA AND
: TITLE OF INVENTION: THEIR USE AS PROBES OR PRIMERS
: NUMBER OF SEQUENCES: 53
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: OLIFF & BERRIDGE
: STREET: P.O. BOX 19928
: CITY: ALEXANDRIA
: STATE: VA
: COUNTRY: USA
: ZIP: 22320
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/632,470
: FILING DATE: 08-JUL-1996
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: BERRIDGE, WILLIAM P
: REGISTRATION NUMBER: 30,024
: REFERENCE/DOCKET NUMBER: WPB 38238
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703)-836-6400
: TELEFAX: (703)-836-2787
: INFORMATION FOR SEQ ID NO: 53:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1484 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
US-08-632-470-53

Query Match 52.3%; Score 772.6; DB 2; Length 1484;
Best Local Similarity 74.1%; Pred. No. 5.2e-269;
Matches 1082; Conservative 0; Mismatches 337; Indels 42; Gaps 7;

QY 1 gatcatgctcagaactaacgctggcgccgctcttaaacatgcaagtcgagcggttag 60
DB 15 gattctggctcagattgaaacgctagcggtatgctttaaaccatgcaagtcgagcggttag 74
QY 61 caatacct-----agcggcgaacgggtgagtaacacgtggttaactctt 102
DB 75 CAGGGAGNCTTNCCTCTGCGCGCGAGAGTGGCGGAGTGAATGCGTAGGAATCTA 134
QY 103 cctccgagctctgggataactttccgaaggaaagtaataaccggatagctctgttgatc 162
DB 135 CTTGTAGTGGGGGATAAACCCTGGGGAACCTCGGGCTAATACCCTAATCTCTTTGGAGC 194
QY 163 acaagattgatagtgaaagattattgcttgagatagccgcgcgcgattagtagt 222
DB 195 AAAGCGGGGGAT--CTTCGGACCTCGTGTATTAAGATGAGCCTACGTCGGATTAGCTTGT 252
QY 223 tggtaggtaattgggtcaccaaggagacatcggtagccggcctagaggggtgtccggcc 282
DB 253 TGGTGGGGTAATGGCTACCAAGGCGACCATCGTAGCTGGTCTGAGAGGACGATCAGCC 312
QY 283 acaatgaactagacacagtgctcactactcctacggaggagcagcagtaagaatctgtc 342
DB 313 ACATGGGATGAGACNCGGCGCAGACTCCTACGGGAGGACGATGGGGAATATTGGAC 372
QY 343 aatgggggaacccctgaacgagcgcgcgtgaacgaaggctcttcgattgttaa 402

DB 373 AATGGGGGAAACCCCTGATCCAGCAATGCCGCGTGTGTGAAGAAGCCCTTCGGGTGTAAA 432
QY 403 gttcattagcaggaataaagcagcaatgtga-----tgatggttacctgc 449
DB 433 GCACCTTCGGTGGGAGAGAAATCTCAAGGGTAATATCTCTNGGCGGTTGACGTTACCCAC 492
QY 450 cta--aacgacgggtaactacgtgcccgcgcgcgtaataacgtatggtgaagcgtt 507
DB 493 AGAAGAAGCACTGGCTAACTCTGTGCCAGCAGCCGCGGTAAATACAGAGAGTCAACGCGTT 552
QY 508 gttcgaataatcattggcgtaaaagggtcgtaggcggaattgttaagtcagggtgtgaaaa 567
DB 553 AATCGGAATCACTGGCGGTAAAGCGCGTGGTATTTAAATCGGATCGAAGGCC 612
QY 568 gcgggctcaacccgtggcctgcacttgaactacaagctctggagtttgggagagcaagt 627
DB 613 CTGGGCTNNNCTGGGAATTGCACCCGATCTTGGGTATCTTGGATATGATAGGGAANT 672
QY 628 ggaattccagggtgtagcggtgaaatcgtagatatacttgaggagaacacccagtggaagcc 687
DB 673 GGAATTCGGGTGTAGCGGTGAATGCGTAGATATCGGAAGAACACCAAGTGGCGAAGCC 732
QY 688 gacttgcctggtcctcaaaactgacgctgagcgcaagcggtggtggtggtggtggtggtggt 747
DB 733 GACTTCCTGGACCAATACTGACACTGAGCGCGCAAGCGTGGGGAGCAACAGGATTAGA 792
QY 748 taccgcggtatccacgccttaaacgttgtctaccagttgttgggggttttaacccctcag 807
DB 793 GACCTGGTGTAGTCCACGCGNGTCAAGGATGAGAACTAGCTGTTGGGAAGTTCCTCTTTAG 852
QY 808 taacgaactaacggattaaagtagacccgctggaactatgctcgaagagtgaaactca 867
DB 853 TAGCGAAGCTAACCGGTTAAGTTCTCCGCTGGGAGTACGGCCGCAAGGTTAAACTCA 912
QY 868 aaggattgacgggggtccgcacaagcggtgagcgatgtggttttaattcgtatgataccccc 927
DB 913 AAGAAATTCACGGGGCGCGCACAAAGCGGTGAGCATGTGGTTTAAATTCGATGCAACGGC 972
QY 928 aaaaacctcaacctggggtgacatggtatgtaacatgtagagatatagacgtcttcgg 987
DB 973 AAAAACTTACCTACCTTACATCTCTCGGAATCTGTGAGAGATGATTTGGTGCCTTCGG 1032
QY 988 ----caattacaggtgctgaggtgtgctcagctcgtgctgagagtggtgggttaa 1044
DB 1033 GAACCGAGTGAAGGTGTGATGCGTGTGAGTGTGCGTGTGAGATGTGGGTAA 1092
QY 1045 gtccgcgaacgagcgcaacccctatcgtatgtgttacc---ttaagttggcactcggtta 1101
DB 1093 GTCCCGTAACGAGCGCAACCCCTCGTCTTAGTTGCCAGCGAGTCAAGTCGGGAACCTCAA 1152
QY 1102 cgaactgcccgtgacaaacccggaggagcggtggtggtggtggtggtggtggtggtggtggt 1161
DB 1153 GGAGACTGCGGTGTATAAACCGGAGGAGGTGGGATGATGTCAAGTCATCATGCGCCCTT 1212
QY 1162 atgtccagggccacacacgctgctacaatggcgatagcagaggtcgcccaactcgcaagag 1221
DB 1213 ACGGTAGGGCTACACACGCTGTACAAATGGGAGTACAAAGGGGTTCGAAGCGCGGAGGT 1272
QY 1222 ggaacttaattctctaaagtccggtcccgatggtgggtggtggtggtggtggtggtggtggt 1281
DB 1273 GGAGCTAATCCCAAGAAACTGCTGTGTAGTCCGGATTGGAGTGTGCAACTCGACTCCATGA 1332
QY 1282 agtcggaatcgctagtaactcgcggatcagcagtcgcgcggtgtaatacgttcccggaacttg 1341
DB 1333 AGTTGGAATCGGTAGTAATCGGGAATCAGCATGTCTCGCGGTGAATACGTTCTCGGGCCTTG 1392
QY 1342 tacacaccccgctcac 1401
DB 1393 TACACACCCCGCTCACACCATGGGAGTGAATGTGTACCAAGAGCGG-GTAGGGCTAACCTT 1451
QY 1402 aaggagacagactactaaggt 1422
DB 1452 CGGGAGCGCGCTCACACACCGT 1472


```

Qy 1407 gacagactactaaggtgaactcgtataaggggggtgaagtcgtatacaaggtacogtaaatc 1466
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1452 GCAGCGUUAACCGAGGAGUGAUCAUGNNNNNNNNNNNNNNNNNNNGUAAACAAGNNNNNNNN 1511
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1467 ga 1468
      | |
Db 1512 GA 1513

RESULT 14
US-08-299-810A-27
; Sequence 27, Application US/08299810A
; Patent No. 5721097
; GENERAL INFORMATION:
; APPLICANT: Rossau, Rudi
; APPLICANT: Van Heuverswyn, Hugo
; TITLE OF INVENTION: HYBRIDIZATION PROBES FOR THE
; TITLE OF INVENTION: DETECTION OF BRANHAMELLA CATARRHALIS STRAINS
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant & Gould
; STREET: 3100 No. 5721097west Center
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/299,810A
; FILING DATE: 01-SEP-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hillson, Randall A.
; REGISTRATION NUMBER: 31,838
; REFERENCE/DOCKET NUMBER: 8076.70-US-WO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-332-5300
; TELEFAX: 612-332-9081
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1485 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Branhamella catarrhalis
; IMMEDIATE SOURCE:
; CLONE: 16S rRNA Gene
US-08-299-810A-27

```

Qy	231	taatggtctacacaagcgacgatcgyttagtcggccgttgagagggtgtcccggccacaatgga	290
Db	208	TAAAGGCGCTACCAAGCGCAGCATCTGTACTGTCTGAGAGATGATCAGCCACACTGGG	267
Qy	291	actagaacacaggtccatactccacgaaagcacagattaaqaacttgtcaatggggg	350
Db	268	ACTGAGACACGGCCCCAGACTCTCTACGGGAGGCAGACGTGGGAATATTGGACAANTGGCG	327
Qy	351	aaacctgaagcagcgacgcgcgtgaaacgaagaagttcttcggattgttaaagtctcat	410
Db	328	NAAGCCTGATCCAGCCATGCCGCGTGTGTGAAGAAGGCCCTTTTGTTGTTAAGCACCTTA	387
Qy	411	ggcagaaaaataaagcagcaaatgtgatgatggttacctgccta-----aag	455
Db	388	AGTGGGGAGGAAAAGCTTATGTTAATACCCATAAGCCCTGACTGTACCACAGAATAAG	447
Qy	456	caccgctlaactcacgtgccagcagccgcgtaactacgtatgatacgtgaacgctgttcgaaa	515
Db	448	CACCGGCTTAACCTGTGTGCCAGCAGCCCGGTAAATACAGAGGGTGTCAAGCGTTAATCGG-A	506
Qy	516	tcatitggcgctaaagggtgcgtagcgcgatttgttaagtcagggtgtaaaactgcgggctc	575
Db	507	TTACTGGGCGTAAGCGCGGTAGTGGTTATTAAAGTCAGATGTGAAGCCCGCGGCTT	566
Qy	576	aacctgtggcctgcacttgaaactacaagctcgtggatttgggagagcgcaagtgaattcc	635
Db	567	AACCTGGGAACCTGCATCTGATACTGGATAAECTAGACTAGTAGGTGAGAGGGNGTAGAATTCC	626
Qy	636	agggttagcgggtgaatcgcgtagatatctctggaagaaacaccagtcgcgaaggcgacttgc	695
Db	627	AGGTGTAGCGGTGAATCGGTAGAGATCTGTGAGGAATACCGATGTGCCGAAGGCAAGCTCCCT	686
Qy	696	ggctcaaacactgacgctgagcgcaagaacgctgggttagtaaaacgggattagataccgccg	755
Db	687	GGCATCATACTGACACTGAGGTGCCAAGACGCTGGGTAGCAACACGAGTTAGATACCCTGG	746
Qy	756	taatccacgcgcctaaacgttgtctaccagttgttgggggtttttaacc-ctcagtaacgaa	814
Db	747	TAGTCCACGCGCTAAACAGATGCTACCACTGCTTTGGGTCTTTTAAAGACTTTAGTGACGCA	806
Qy	815	cctaacggattaaagtagaccgcctgggactatgctcgcgaagagtgaaactcaaaagaaat	874
Db	807	GTTAACGCAATAAGTAGACCGCTGGGGAGTAGCGGCCGAAGGTTAAAACTCAAATGAAT	866
Qy	875	tgacgggggtccgcacaagcggtaggacatgtggtttaaattcgatgatcccccaaaacc	934
Db	867	TCACGGGGCCGCACACGCGGTGAGCATGTGGTTAATTTCGATGCACACGCCGAAGACC	926
Qy	935	tcacctgggcttgacatggatcgaatcattagagatatatgagccttcgggcagattc	994
Db	927	TTACCTGGCTTTGACATAG-TGAGAATCTTGCAGAGATGCGAGAGTGCCTTCGGGAATTC	985
Qy	995	---acaggtgctgatggttgtctcaagctcgttgatgaatgttggtttaagtcccg	1050
Db	986	ACATACAGGTGCTGATGGCTCTCGTCACTCGTGTGAGATGTGTGGGTTAAGTCCC	1045
Qy	1051	caacgagcgcaacccttatgttgttacc--ttaagttgggcacttgggttaagtcgaaact	1108
Db	1046	CNACGAGCGCAACCTTTTCTTAGTTACACGGACTCGTTCGGGAACCTCTAGGATACT	1105
Qy	1109	gccggtgcaaaacccggaggaagcggggtagcgtcaaatcctctcatggcctttatgtcca	1168
Db	1106	GCCAGTGACAACTGGAGAAAGCGGGGACGAGCTCAAGTCATCATGGCCCTTACGACCA	1165
Qy	1169	ggggcacacacagtgctacaaatggccgatatcacaggggtcgcccaactcgcaagaggsgcta	1228
Db	1166	GGGCTATACACACGTGTACAATGGTTGGTACAAGGGTGTCTACACAGCGATGTGATGCTA	1225
Qy	1229	atctctaaaagtcggtlcccgattcggattggggttcgaactcgaccccatgaagtcgga	1288
Db	1226	ATCTCAAAAAGCCAATCGTAGTCCGGATTGGAGTCTTGCAACTCGACTCCTCATGATGAAGTCGGA	1285
Qy	1289	atcgcgtagtaatcgcggatcagcatgcgcgcggtgaaatcagttcccgagccttgtacacac	1348

```
|||||
Db 1286 ATCCCTAGTAATCGCAGATCAGAAATGCTGCGGTGAATACGTTCCGGGCCCTGTACACAC 1345
Qy 1349 cccccccacacccctgagtggggagcaccggaagtgtcttggtaaccgtaaggaga 1408
Db 1346 CGCCGTCACACCATGGGAGTGTATCTCACCAGAAGTGTGTAGCCTAACGC---AAGAGG 1402
Qy 1409 cagactactaagggtgaactcgtataaagggtgaagtgcgtacaaggtacc 1459
Db 1403 GCGATCACACGGTGGGTGTCGATGACTGCTGGGTGAAGTGTGAACAGGTAGC 1453

RESULT 15
US-08-875-445-21
; Sequence 21, Application US/08875445
; Patent No. 5969642
; GENERAL INFORMATION:
; APPLICANT: Kanta SAKAMOTO
; TITLE OF INVENTION: DETECTION OF THE GENUS PECTINATUS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 kb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/875.445
; FILING DATE: July 28, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lee Cheng
; REGISTRATION NUMBER: 40,949
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX: 202-371-8856
; TELEX:
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1542
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; ORIGINAL SOURCE:
; BIOLOGICAL NAME: Pectinatus
; ORIGINAL SOURCE: Pectinatus sp. DSM 20764
US-08-875-445-21

Query Match 51.9%; Score 766.6; DB 2; Length 1542;
Best Local Similarity 74.7%; Pred. No. 7.9e-267;
Matches 1126; Conservative 0; Mismatches 329; Indels 52; Gaps 11;

Qy 1 gatcgtgctcagaactaacgtgcgcgctgtcttaaacatgcaagtgcagcggttag 60
Db 8 GATCTGCTCAGGACGAACGCTGCGGCGTCTTAACACATGCAATGCAAGTCGAACGGGACTT 67
Qy 61 caatac-----ctagcgcggaacgggtgagtaacacgt-ggtaattcttcctcga 109
Db 68 TTATTTCCGTAATAAGTCTAGTGGCAACCGGTGAGTAACCGCTAGCGCAACCTTACCTCAA 127
Qy 110 gtcctggataacttccgaagggaagtaataaccggatagtcctgttggatcacagaat 169
Db 128 GATGGGGACAACATCCCGAAAGGGGTGCTTAATACCGAATGTTGTAAAGAGTACTGCATGCT 187
```

```
Qy 170 ttgataggtaaag-----atttattgttggagatgagcccgccgagcattagcta 220
Db 188 ACTTTTACAAAACGGCGCTTTTACTGTGTACTGTAGATGGCGTCTGCTCTCTAGTAGTA 247
Qy 221 gttgtga-ggtaatggctcaccgaagcgacgtagcggtagccgctcagaggggtccg 279
Db 248 GTTGTGACGGTAAATGGCGCACCAAGGCAACGATCAGTAGCCGGTCTGAGAGATGGACG 307
Qy 280 gccacaatggaaactgagacacggttccatcactcctacggagcgagcagcttaagaactctg 339
Db 308 GCCACATTGGGACTTGAGACACCGGCCAGACTCCTACGGAGGCGACGACACTGGGAATCTTC 367
Qy 340 ctcaatgggggaaacccctgaagcagcgagcgcggtgaacgaagaaggtctctcgatagt 399
Db 368 CGCAATGGCGCAAAAGCCTTGACGGAGCAACCGCGGTGAACGAGGAAGTCTTCGGATCGT 427
Qy 400 aaagtctcataggcaggaataaagaacgaatgtagatgggtacctgctct----- 451
Db 428 AAAGTCTCTTGCAGGGGACCAATGGCATTTAGTCTAATACCACTAATGAATGACGGTAC 487
Qy 452 -----aaagcaccggcttaactacgtgcccagcagccggttaatacgtatggtgcag 503
Db 488 CTTGTTAGAAAACGACGGCTAACTACGTGCCAGCAGCGCGGTAAATACGTAGCGCGCAAG 547
Qy 504 cgtgttcggaatcatctggcgtaagggtgtagtagcggtattgttaagtcaggtgtgaa 563
Db 548 CGTTGTCGGAAATCATTTGGCGTAAAGGAGCGCAGCGGACATTTAAG-CGGATCTTTAA 606
Qy 564 aactgcgggctcaaccgtgcccgtcacttgaacctcaagtctgaggtttgggagagcg 623
Db 607 AAGTGGGGGCTCAACCCCGTGCATGGGGTCCGAACCTGAATGTCTGTAGTGCAGGAGAGA 666
Qy 624 aagtggaaattccaggtgtagcgggtgaaatgcgtagatcatctggagggaacacagtg 683
Db 667 AAGCGGAATTCCTAGTGTAGCGGTGAAATGCGTAGATATTGGGAAGAACACACGATGGGA 726
Qy 684 aggcacttctgctcaaaactaacgctgagcgcaaaagcgtggttagtaaaacggat 743
Db 727 AGCGGCTTCTGGACTGTACTGACGTGAGGCTCGAAGCCAGGCTAGCGAAGCGGAT 786
Qy 744 tagatcccccggttaatccacgccttaacgttgtctacagttgttgggggttttaa--c 801
Db 787 TAGATACCCCGTAGTCTCTGCGCGTAAACGATGGATACTAGGTGTAGGGGTATCGACCC 846
Qy 802 cctcagtaacgaacctaacggattaaagtagaccgctctgggagactatgctcgaagagta 861
Db 847 CCCCTGTGCGGAGTTAACGCAATAAGTATCCCGCTGGGAGTACGCGCCGCAAGGCTGA 906
Qy 862 aactcaaggaattgacgggggtccgcacaagcggtgagcatgtggttaattcgtatga 921
Db 907 AACTCAAGAGGAATTGACGGGGGCCCGCACACGCGTGGAGTATGTGTTAATTTCGACGC 966
Qy 922 taccocaaaaacctcacctgggcttgacatggatcgtgaatcatgtagagatatagacc 981
Db 967 AACGCGAAGAACTTTACCAGGGCTTTGACATTGAT-TGACGCAATTCAGAGATGGATGCTTC 1025
Qy 982 ttc-----gggcagattcagaggtgctgcatggtgctcagctcgtgctgtagatg 1035
Db 1026 CTCTTCGAGGAGCAAGAAACAGAGTGGTGCATGGTGTGCTGCTGAGTGTGCTGAGATG 1085
Qy 1036 ttgggttaagtcctccgaacgacgaacccctctctgtatgtgc---taccttaagtgg 1092
Db 1086 TTGGGTTAAGTCCCGCAACGAGCGCAACCCCTATCATTTGTGCCAGCACGCTAACCGGTG 1145
Qy 1093 gcactgggtlacgaaactgcgggtgacaaaacggaggaagcggtgagcgtcaaatcctc 1152
Db 1146 GAACTCAATGAGACTGCGCGGACACACGCGGAGGAGCGGGGATGACGTCAAGTCATC 1205
Qy 1153 atggccttatgtccagggccaacacgctgctacaatggccgatacagaggttcgccaac 1212
Db 1206 ATGCCCTTACGTCTCTGGGCTACACACGTACTACAAATGGGATACACAGAGGAAGCAAG 1265
Qy 1213 tcgcaagggggagcctaactctctaaaaagtcgggtcccgatggtgggtctgcgaactcg 1272
```

Db	1266	GAGGATTCGGAGCGGAACCCAAAAAATATCCCACAGTTCGGATTTCAGACTGCAACTCG	1325
Qy	1273	accccatgaagtcggaatcgtagtaatcgcgcatcagcatcgcgcggtgaaatcaccttcc	1332
Db	1326	CCTCATGAAGTCGGAATCGCTAGTAAATCGCAGGTCAGCATACTCGCGTCAATACGTTCC	1385
Qy	1333	cggaccttgtacacacgcgcgtcacaccactcgtaggggagcaccccgagtggtcttt	1392
Db	1386	CGGCGCTTGTACACACCGCCCGTCACACCGAAGTCATTTCACCCCGGAAGCGCGCTAA	1445
Qy	1393	gttaacggtaaagagacagactactaaagtgaaactcgtaaaggggtcgaagtcgtaaca	1452
Db	1446	G--GGCTTATGGAACCGACCGACCTCTAAGGTGGGGCGATGTTGGGTGAAGTCGTAACA	1503
Qy	1453	aggtacc	1459
Db	1504	AGGTAGC	1510

Search completed: March 27, 2001, 08:19:18
Job time: 5145 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 27, 2001, 06:21:57 ; Search time 2517.78 Seconds
(without alignments)
4110.788 Million cell updates/sec

Title: US-09-380-826A-1
Perfect score: 1477
Sequence: 1 gatcatggctcagaactaac.....ccgtaaatcgattcctgcag 1477.

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues

Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:
1: gb_est1.*
2: gb_est2.*
3: gb_est3.*
4: gb_est4.*
5: gb_est5.*
6: gb_est6.*
7: gb_est7.*
8: gb_est8.*
9: gb_est9.*
10: gb_est10.*
11: gb_est11.*
12: gb_est12.*
13: gb_est13.*
14: gb_est14.*
15: gb_est15.*
16: gb_est16.*
17: gb_est17.*
18: gb_est18.*
19: gb_est19.*
20: gb_est20.*
21: gb_est21.*
22: gb_est22.*
23: gb_est23.*
24: gb_est24.*
25: gb_est25.*
26: gb_est26.*
27: gb_est27.*
28: gb_est28.*
29: gb_est29.*
30: gb_est30.*
31: gb_est31.*
32: gb_est32.*
33: gb_est33.*
34: gb_est34.*
35: gb_est35.*
36: gb_est36.*
37: gb_est37.*
38: gb_est38.*
39: gb_est39.*
40: gb_est40.*
41: em_estba.*
42: em_estfun.*
43: em_esthum1.*

44: em_esthum2.*
45: em_esthum3.*
46: em_esthum4.*
47: em_esthum5.*
48: em_esthum6.*
49: em_esthum7.*
50: em_esthum8.*
51: em_esthum9.*
52: em_esthum10.*
53: em_esthum11.*
54: em_esthum12.*
55: em_esthum13.*
56: em_esthum14.*
57: em_esthum15.*
58: em_esthum16.*
59: em_esthum17.*
60: em_esthum18.*
61: em_esthum19.*
62: em_esthum20.*
63: em_estin1.*
64: em_estin2.*
65: em_estin3.*
66: em_estin4.*
67: em_estov1.*
68: em_estov2.*
69: em_estpl1.*
70: em_estpl2.*
71: em_estpl3.*
72: em_estpl4.*
73: em_estpl5.*
74: em_estro1.*
75: em_estro2.*
76: em_estro3.*
77: em_estro4.*
78: em_estro5.*
79: em_estro6.*
80: em_estro7.*
81: em_estro8.*
82: em_estro9.*
83: em_estro10.*
84: em_estro11.*
85: em_estro12.*
86: em_estro13.*
87: gb_est41.*
88: gb_est42.*
89: gb_est43.*
90: gb_est44.*
91: gb_est45.*
92: gb_est46.*
93: gb_est47.*
94: gb_est48.*
95: gb_est49.*
96: gb_est50.*
97: gb_est51.*
98: gb_est52.*
99: gb_est53.*
100: gb_est54.*
101: gb_est55.*
102: gb_est56.*
103: gb_est57.*
104: gb_est67.*
105: gb_est68.*
106: gb_est69.*
107: gb_est70.*
108: gb_est71.*
109: gb_est72.*
110: gb_est73.*
111: gb_est74.*
112: em_esthum21.*
113: em_esthum22.*
114: em_esthum23.*
115: em_estom1.*
116: em_estom2.*

117: em_estpl6:*
 118: em_estpl7:*
 119: em_estpl8:*
 120: em_estrol4:*
 121: em_estrol5:*
 122: em_estrol6:*
 123: em_estrol7:*
 124: em_estrol8:*
 125: em_estrol9:*
 126: gb_est58:*
 127: gb_est59:*
 128: gb_est60:*
 129: gb_est61:*
 130: gb_est62:*
 131: gb_est63:*
 132: gb_est64:*
 133: gb_est65:*
 134: gb_est66:*
 135: gb_est75:*
 136: gb_est76:*
 137: gb_est77:*
 138: gb_est78:*
 139: gb_est79:*
 140: gb_est80:*
 141: gb_est81:*
 142: gb_est82:*
 143: gb_est83:*
 144: gb_est84:*
 145: gb_est85:*
 146: gb_est86:*
 147: gb_est87:*
 148: gb_gss1:*
 149: gb_gss2:*
 150: gb_gss3:*
 151: gb_gss4:*
 152: em_gss1:*
 153: em_gss2:*
 154: em_gss3:*
 155: em_gss4:*
 156: gb_gss5:*
 157: gb_gss6:*
 158: gb_gss7:*
 159: gb_gss8:*
 160: gb_gss9:*
 161: gb_gss10:*
 162: gb_gss11:*
 163: gb_gss12:*
 164: gb_gss13:*
 165: gb_gss14:*
 166: gb_gss15:*
 167: gb_gss16:*
 168: gb_gss17:*
 169: gb_gss18:*
 170: gb_gss19:*
 171: gb_gss20:*
 172: gb_gss21:*
 173: gb_gss22:*
 174: gb_gss23:*
 175: gb_gss24:*
 176: em_gss5:*
 177: em_gss6:*
 178: em_gss7:*
 179: em_gss8:*
 180: em_gss9:*
 181: em_gss10:*
 182: em_gss11:*
 183: em_gss12:*
 184: em_gss13:*
 185: em_gss14:*
 186: em_gss15:*
 187: em_gss16:*
 188: em_gss17:*
 189: em_gss18:*

190: gb_gss25:*
 191: gb_gss26:*
 192: gb_gss27:*
 193: gb_gss28:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	436.8	29.6	800	165	AQ960400	LERFD51TR
C 2	424	28.7	716	165	AQ957363	LERAP36TR
C 3	407.4	27.6	720	190	CNS000CV	AL093421 Arabidops
C 4	389	26.3	746	165	AQ957362	LERAP36TF
C 5	383.6	26.0	741	165	AQ957181	LERAO33TR
C 6	378.8	25.6	757	165	AQ957182	AQ957182 LERAO34TR
C 7	376.6	25.5	676	175	B73824	B73824 T27K4TR TAM
C 8	360.4	24.4	693	165	AQ956298	AQ956298 LERAJ19TR
C 9	359.6	24.3	676	165	AQ969911	AQ969911 LERJR28TF
C 10	356.6	24.1	659	165	AQ969996	AQ969996 LERJR90TF
C 11	352.4	23.9	677	87	AW221851	AW221851 EST298662
C 12	350.6	23.7	616	190	AG019017	AG019017 Homo sapi
C 13	346.8	23.5	675	165	AQ956297	AQ956297 LERAJ19TF
C 14	344	23.3	659	95	AW833978	AW833978 QVO-T1000
C 15	343	23.2	603	37	AV531914	AV531914 AV531914
C 16	340.4	23.0	768	165	AQ957272	AQ957272 LERAO84TR
C 17	340	23.0	637	165	AQ962145	AQ962145 LERGD59TR
C 18	339.6	23.0	687	165	AQ955848	AQ955848 LERAG49TR
C 19	339	23.0	632	165	AQ967817	AQ967817 LERIV15TF
C 20	337.4	22.8	618	165	AQ969912	AQ969912 LERJR28TR
C 21	336.6	22.8	601	174	B27166	B27166 T2C12TR TAM
C 22	334.2	22.6	660	92	AW600903	AW600903 3B5 CDNA
C 23	333.6	22.6	592	190	AG019018	AG019018 Homo sapi
C 24	332.4	22.5	599	165	AQ968703	AQ968703 LERJE31TF
C 25	332	22.5	648	165	AQ957342	AQ957342 LERAP25TF
C 26	331.6	22.5	561	190	AG019935	AG019935 Homo sapi
C 27	331.6	22.5	616	107	BE356988	BE356988 DGL_146_B
C 28	331.6	22.5	673	148	AF073762	AF073762 AF075762
C 29	330.8	22.4	604	105	BE204135	BE204135 EST396811
C 30	327.6	22.2	662	94	AW736167	AW736167 EST332153
C 31	327.2	22.2	639	165	AQ956183	AQ956183 LERAI55TR
C 32	324.4	22.0	610	165	AQ961463	AQ961463 LERPL40TR
C 33	319.8	21.7	553	107	BE362383	BE362383 DGL_86_H0
C 34	319	21.6	649	93	AW675839	AW675839 SWYD25CAU
C 35	315.2	21.3	567	165	AQ961462	AQ961462 LERPL40TF
C 36	314.2	21.3	780	190	CNS000XIN	AL094925 Arabidops
C 37	312	21.1	584	165	AQ964113	AQ964113 LERGT38TF
C 38	311.6	21.1	583	165	AQ967819	AQ967819 LERIV15TR
C 39	310.6	21.0	503	92	AW563518	AW563518 LGL_239_H
C 40	310	21.0	676	105	BE187622	BE187622 EST336183
C 41	308.2	20.9	766	111	BE661742	BE661742 969 Gmax5
C 42	307.4	20.8	714	105	BE202334	BE202334 SVOVMFCAR
C 43	304.4	20.6	540	165	AQ968704	AQ968704 LERJE31TR
C 44	302.4	20.5	899	174	B10095	B10095 F8G13-T7 IG
C 45	299.4	20.3	553	165	AQ966081	AQ966081 LERII07TF

ALIGNMENTS

RESULT 1
 AQ960400/c
 LOCUS
 DEFINITION LERFD51TR LERA Arabidopsis thaliana genomic clone LERFD51, DNA
 800 bp
 DNA
 28-JAN-2000
 ACCESSION AQ960400
 VERSION AQ960400.1
 GI:6788101
 KEYWORDS
 GSS.
 thale cress.

ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE 1 (bases 1 to 800)
 AUTHORS Buell C.R., Lin X., Pai G., Barnstead, M., Bowman, C., Utterbach, T.,
 Feldblyum, T., Liang, F., Creasy, T. and Fraser, C.M.
 TITLE Genomic survey sequencing of Landsberg erecta ecotype of
 Arabidopsis thaliana and identification of sequence-based
 polymorphisms
 JOURNAL Unpublished (2000)
 COMMENT Contact: Xiaoying Lin
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: at@tigr.org
 For additional information, see <http://www.tigr.org/tdb/at/at.html>
 Similar to A. thaliana chloroplast sequence (GB:AF000423)
 Seq primer: TR
 Class: shotgun.

FEATURES

source
 Location/Qualifiers
 1..800
 /organism="Arabidopsis thaliana"
 /strain="Landsberg erecta"
 /db_xref="taxon:3702"
 /clone="LERFD51"
 /clone_lib="LERA"
 /note="Organ: Leaf; Vector: PHOS1; Total genomic DNA was
 sheared to 0.9-1 kbp before ligation."
 147 a 257 c 181 g 215 t

Query Match 29.6%; Score 436.8; DB 165; Length 800;
 Best Local Similarity 74.9%; Pred. No. 1.2e-127;
 Matches 589; Conservative 0; Mismatches 187; Indels 10; Gaps 3;

QY 190 gtttgagatgagccgagcgttagctagttggtgagtaagtgcctcaccaggcga 249

Db 780 GCCCGAGGAGGGCTCGGCTCTGATTAGTGTGGTGGCAATAGCTTACCAAGGCCA 721

QY 250 cgtcgttagcgcgctgaagggtgtccggcccaatggaactagacacggtccatc 309

Db 720 TGATCAGTAGCTGGTCCGAGAGATGATCAGCCACACTGGGACTGAGACCGGCCAGAC 661

QY 310 tctcagggagcagcagtgtaagaatcttctcgaatgggggaaacccctaaagcgcagc 369

Db 660 TCCTACGGGAGGACGACGAGTGGGAATTTTCGCAATGGCGAAGCCTGACGAGCAATG 601

QY 370 ccgctgaaacgaagtgcttcggttgatgtgaagtctattagggcggaataaagcagc 429

Db 600 CCGCGTGGAGGTAGAGGCGCTACGGGCTCTGAACTTTTTCCTCCAG--AGAAGAAGCAAT 543

QY 430 aatgtgatgtgacctgcctaaagcagcgcgttaactacgtgccagacgcgcgtta 489

Db 542 GACGGTATCTGGGA-----ATAAGCATCGCGTAACTCTGTGCCAGCAGCCCGGTAAT 489

QY 490 acgtatgtgcgaagcgtgttctgggaatcattggcggttaaggtgctgagcggtttgt 549

Db 488 ACAGAGGATGCAAGCGTTATCCGGAATGATTGGCGTAAAGCGTCTGTAGTGGCTTTT 429

QY 550 aagtcaggtgtgaaactgcygggtcctaaacccgtgaccttgaanaactacaagtcgtg 609

Db 428 AAGTCGCGCGTCAAAATCCAGCGGCTCAACCCCTGACAGCGCGGTGGAACATACCAAGCTTG 369

QY 610 agtttggagagcgaagtgaattccaggttagcgtgtagcgtgtaaatggttagatctcgtgga 669

Db 368 AGTACGGTATGGCGAGAGGGAATTTCCGCTGAGCGGTGAATGCGCTAGAGATCGGAAG 309

QY 670 aacaccagtcgcgaagcgactgctggtctcaaaactgacgtgaggcacgaagcgttg 729

Db 308 AACACCAACGGCGAAGCACTCTGCTGGCCGACACTGACACTGAGACGACGAAGCTAGG 249

QY 730 gtagtaaacgggattagataccgccgtaatccacgccctaaacgttgttaccagttgtt 789

Db 248 GGAGCGAATGGGATTAGATACCCCACTAGCTCTAGCCGTAAACGATGGATACTAGCGCCT 189

QY 790 ggggttttaaccc--tcagtaacgaacccaacgattaaagtacacccgctgggaactat 847

Db 188 GTGCGTATCGACCCGTCAGTCTGTAGCTAACGGCTTAAGTATCCCGCTGGGAGTAC 129

QY 848 gtcgcaagagtgaactcaaggaattacggtgggtccgcacaaagcgttgagcatgtg 907

Db 128 GTTCGCAAGATGAACACTCAAGGAATTCACGGGGCCGCGACACAGCGGTGAGCATGTG 69

QY 908 gtttaattogatacccccacccctcacctgggttgatgacatgatatcatgta 967

Db 68 GTTTAATTCGATGCAAGGCGAAGAACCTTTACCAGGGCTTGACATGCCGGAATCCTCTTA 9

QY 968 gagata 973

Db 8 AAGAGA 3

RESULT

2

AO957363/c

LOCUS

AO957363

DEFINITION

AO957363

ACCESSION

AO957363

VERSION

AO957363.1

KEYWORDS

GSS.

SOURCE

thale cress.

ORGANISM

Arabidopsis thaliana

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;

Brassicales; Brassicaceae; Arabidopsis.

REFERENCE

1 (bases 1 to 716)

AUTHORS

Buell C.R., Lin X., Pai G., Barnstead, M., Bowman, C., Utterbach, T.,

Feldblyum, T., Liang, F., Creasy, T. and Fraser, C.M.

TITLE

Genomic survey sequencing of Landsberg erecta ecotype of

Arabidopsis thaliana and identification of sequence-based

polymorphisms

JOURNAL

Unpublished (2000)

COMMENT

Contact: Xiaoying Lin

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: at@tigr.org

For additional information, see <http://www.tigr.org/tdb/at/at.html>

Seq primer: TR

Class: shotgun.

Location/Qualifiers

1..716

/organism="Arabidopsis thaliana"

/strain="Landsberg erecta"

/db_xref="taxon:3702"

/clone="LERAP36"

/clone_lib="LERA"

/note="Organ: Leaf; Vector: PHOS1; Total genomic DNA was

sheared to 0.9-1 kbp before ligation."

BASE COUNT

144 a 213 c 154 g 205 t

ORIGIN

Query Match 28.7%; Score 424; DB 165; Length 716;

Best Local Similarity 77.6%; Pred. No. 1.3e-123;

Matches 545; Conservative 0; Mismatches 140; Indels 17; Gaps 2;

QY 275 gtccggccacaaatggaactgagacaggtccatactcctacgagcgagcagcagtaagaa 334

Db 716 GACCACACCATGGGAACCTAAGACACGGTCCAGACTCTTACGGGAGCGACAGCTGGGAA 657

QY 335 tcttgctcaatgggggaaacccctgaagcagcagcgcgcgtgaaacgaagaagctcttcg 394

```

Db 656 TTTTTCGCAATGGCGCAAGCCTGATCCAGCCATTCGCGTGAGTGAAGAAGCGCTTCGG 597
Qy atttaagtattcattagccagcaaaataagcagcaaatgtgatgtacctcctccta-- 452
Db 596 GTTGTAAAGCTCTTTCCGCAAGGGGAAGAAACTTACTTTCTAATAAAGAGTGAAGCTGACG 537
Qy -----aagcaccggctaaactacgtgcagcagccggtgtaatacgtatggtg 499
Db 536 GTACCTTGATAGAAGCACCAGCTAACCTACGTGCAGCAGCCGGTAAATACGTAGGGTG 477
Qy caagcgttctcgaatacatctggcgttaaaaggtgcgtagcggatttgaagtcaggtg 559
Db 476 CGACCGTTTAATCGCAATTTACTGGCGTAAAGCGTTCGCGAGCGGTTTGGCAAGTCAGATG 417
Qy tgaaaactgcgggctcaaccgtggcctgcactgaaactcaagctcgaagtttggag 619
Db 416 TGAATCCCCGAGCTCACTTGGGAACCTGCGTTTGAACACTGCCAGACTAGATATGTCAG 357
Qy agcaagtggattccaggttagcgtggaatcgtagatatctggagggaacccagtg 679
Db 356 AGGGGGTGAAGATTCACGCTAGCAGTGAATGCGTAGAGATGTGGAGGAATACCAATG 297
Qy gcgaagcgacttgctggtcctcaaaactgaactgagcagcaaaagcgtggtagtaaacg 739
Db 296 GCGAAGCGAGCCCCCTGGGATAATATTGACGCTCATGCACGAAGCGTGGGAGCAACA 237
Qy ggattagatacccggttaataccagccctaaacgtgtctaccagttgttggggtttta 799
Db 236 GGATTAGATACCTTGGTAGTCCACGCCCTAAACGATGCTACTAGTTGTTGGAGTAA 177
Qy 800 a---cctcagtaacaaactaacggattaaagtagacagccctggggaactatctcgaaga 857
Db 176 AATCATGAGTAACGACACTAACCGCTGAAGTAGACCGCCCTGGGAGTAGACGTCGCAAGA 117
Qy 858 gtgaactcaagaagattgacgggtccgcacaaagcgtgagcattgttttaattcg 917
Db 116 TTAAAACTCAAAAGGAATTACCGGGGGCCCCGACAAACCGGTGGATTATGTGATTAAATTCG 57
Qy 918 atgatacccaaaacccctcactggcttgacatggatctga 959
Db 56 ATGCAACGCAAAACCTTACCTGGCCCTTGACATGCCACTAA 15

RESULT 3
CNS00MCV 720 bp DNA GSS 28-JUN-1999
LOCUS Arabidopsis thaliana genome survey sequence Sp6 end of BAC T11H14
DEFINITION of TAMU library from strain Columbia of Arabidopsis thaliana,
genomic survey sequence.
ACCESSION AL093421 GI:5294575
VERSION AL093421.1
KEYWORDS GSS.
SOURCE Salanoubat,M., Choinsne,N., Artiguenave,F., Brottier,P., Wincker,P.,
Sanson,D., Saurin,W., Weissenbach,J. and Quetier,F.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; euroids II; Brassicales; Brassicaceae;
Arabidopsis.
REFERENCE 1 (bases 1 to 720)
AUTHORS Salanoubat,M., Choinsne,N., Artiguenave,F., Brottier,P., Wincker,P.,
Sanson,D., Saurin,W., Weissenbach,J. and Quetier,F.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 720)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (25-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr
FEATURES
Location/Qualifiers
source
1..720
/organism="Arabidopsis thaliana"
/strain="Columbia"

```

```

/db_xref="taxon:3702"
/clone_lib="TAMU"
/clone="T11H14"
/notes="end : Sp6"
BASE COUNT 178 a 175 c 230 g 137 t
ORIGIN

Query Match. 27.6% Score 407.4; DB 190; Length 720;
Best Local Similarity 77.2%; Pred. No. 2.5e-118;
Matches 547; Conservative 0; Mismatches 156; Indels 6; Gaps 4;

Qy 609 gagttgggagaggaagtgaattccagggtgtagcggtagaattcgcgtagatattcgtgag 668
Db 1 GAGTACGGTTGGGCGAGAGGGAATTTCCGGTGAGGCGGTGAATTCGCTAGAGATCGGAAA 60
Qy 669 gaacacagtgccgaagcgaacttgctcgaactcaaacgacgtgagcagcaaaagcgtg 728
Db 61 GAACACCAACCGCGAAGACACTCTGCTGGGCCGACACTGACACTGAGAGAGCAAGAGCTAG 120
Qy 729 ggtagtaaacgggattagataaccgccggttaatccacgcgcctaaacggttctaccagttgt 788
Db 121 GGGAGCGAATGGGATTAGATACCCAGTAGTCTTAGCCGTAACAGATGAGATAGGCGC 180
Qy 789 tgggggttttaaccc--tcagtaaacgaacctaacggattaaagtagacgcgcgtgggacta 846
Db 181 TGTGCGTATCGACCGCGTGACAGTGTGTAGCTTAACGCGTTAAAGTTATCCGCGCTGGGAGTA 240
Qy 847 tgcgcgaagagtgaactcaaaaggaattgacgggggtccgcacaaagcgtgagcagcgtg 906
Db 241 CTTTCGCAAGAAATGAACACTCAAGGAATTTGACGGGGGCGCCGACAAACGCGTGGAGCATGT 300
Qy 907 ggtttaattcgatgatacccccacaaacccctcactcgggtgggttgacatggtatcgaatcgt 966
Db 301 GGTTTAATTTCGATGCAAAAGCGAAGAACCTTACCAGGGCTTGACATGCCGCGAATCCTCTT 360
Qy 967 -agagatatagaccttcggggagat--tcacagggtgctgcatggttgctgagtcgtc 1023
Db 361 GAAAGAGAGGGGTGCCCTTCGGGAACGCGGACACAGGTGTCATGGGTGTCGTGCTGAGCTCG 420
Qy 1024 tgcgtgagatgttgggttaagtccgcgaacgagcgaacccctcactgattgttgcacc 1083
Db 421 TGCGTAAAGTGTGGGTAAAGTCCCGCAACGAGCGCAACCCCTCGTGTAGTTAGTTGCCACC 480
Qy 1084 -ttaagttgggactggttacgaaactgccggtgacaaacccggaaggaagcgggagtgagc 1142
Db 481 GTTGAGTTTGGAAACCTGAACAGACTGCCGGTGATGAAGCCGAGGAGAGGTGAGGATGACG 540
Qy 1143 tcaatcctcatgaccttattgctcaggggccacacacagctgacaaatggtccgatacagag 1202
Db 541 TCAAGTCATCATGCCCTTATGCCCCCTGGGCGACACACGCTGTACAAATGGCGGGGACAAAG 600
Qy 1203 ggtcgcaactcgcaagaggagctaaactcctcaaaagtcggtcccgagtcggattggggt 1262
Db 601 GGTGCGGATCCCGGAGGGTGAGCTAACTCCAAAAACCCGTTCTCAGTTTCGATTGTCAGG 660
Qy 1263 ctgcaactcgccccatgaatcggaatcgctagtaatcgctgagatcagc 1311
Db 661 CTGCAACTCGCTGCATGAAGCGGAATCGTAGTAATATCGCGCGGTACAGC 709

RESULT 4
AQ957362 746 bp DNA GSS 28-JAN-2000
LOCUS LERAP36TF LERA Arabidopsis thaliana genomic clone LERAP36, DNA
DEFINITION sequence.
ACCESSION AQ957362
VERSION AQ957362.1 GI:6785063
KEYWORDS GSS.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; euroids II;

```



```

370 ccgctgaacgaagaaggtcttcgattgttaaagttcattaggcagagaaataaagcagc 429
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 517 CCGCGTGGAGGTAGAAAGCCCTACGGCTCCTCAACTTCTTTTCCAG--AGAAGAAGCAAT 460
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 430 aatgtgatgtgttacctgacctaaagcaccggtctaactacgtgccagcaccggtgtaat 489
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 459 GACGGTATCTGGGA-----ATAAGCATCGGCTAACTCTGTGCCAGCAGCCGCGTAAAT 406
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 490 acgtatgtgcaacgctgtgttcgaatacattggcgtgaaggggtgctagcggatttgt 549
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 405 ACAGAGGATGCAAGCGTTATCCGGAATGATTGGCGGTAAACGCTCTGTAGTGGCTTTT 346
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 550 aagtcaggtgtgaactgcgggtcctaacccgtgacctgacctgaaactacaagcttg 609
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 345 AAGTCCGCGCTCAAAATCCAGGGCTCAACCCCTGCACAGGGCGGTGCAAACTACCAAGCTTG 286
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 610 agtttggagaggaagtggaattccaggttagcgtgaaatcgtagatatctgaag 669
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 285 AGTACGCTAGGGGAGAGGGAATTTCCGGTGGAGCGGTGAATCGGTAGAGATCGGAAG 226
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 670 aacaccagtggcgaagcgacttgtggtctcaaaactgacgctgagcgacgaagcgtgg 729
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 225 AACACCAACGGCGAAGACACTCTCTCGGCCGACACTGACACTGAGAGACGAAAGCTAGG 166
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 730 gtagtaacgggattagataccccggtaatccacgcccctaacgcttctaccagttgt 789
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 165 CGAGCGAATGGATAGATATACCCAGTACTCTAGTCTCAACGCTAAACGATGATAGTGGCGCT 106
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 790 ggggttttaaccc--tcagtaacgaactaagcgtgattaaagtagcgcctgggactat 847
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 105 GTGCGTATCGACCCGTGCGAGTGTGTAGCTAACGCGTTAAAGTATCCCGCCCTGGGAGTAC 46
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 848 gctcgaagtgaaactcaaggaattgacgggggtccgcaca 891
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 45 GTTCCCAAGATGAACATCAAGAGGATTCACGGGGCCCTTACA 2
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 6
ACQ57182/c LERAO34TR LERA Arabidopsis thaliana genomic clone LERAO34, DNA
LOCUS DEFINITION
ACQ57182.1 GI:6784883 GSS
VERSION
KEYWORDS
SOURCE
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
AUTHORS
Buell,C.R., Lin,X., Pal,G., Barnstead,M., Bowman,C., Utterbach,T.,
Feldblyum,T., Liang,F., Creasy,T. and Fraser,C.M.
TITLE
Genomic survey sequencing of Landsberg erecta ecotype of
Arabidopsis thaliana and identification of sequence-based
polymorphisms
JOURNAL
COMMENT
Unpublished (2000)
Contact: Xiaoying Lin
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: at@tigr.org
For additional information, see http://www.tigr.org/tdb/at/at.html
Similar to A. thaliana chloroplast sequence (GB:AF000423)
Seq primer: TR
Class: shotgun.
Location/Qualifiers
1..757
/organism="Arabidopsis thaliana"
/strain="Landsberg erecta"
/db_xref="taxon:3702"
/clone="LERAO34"
FEATURES
source

```

```

/clone_lib="LERA"
/Note="Organ: Leaf; Vector: pHOs1; Total genomic DNA was
sheared to 0.9-1 Kbp before ligation."
BASE COUNT 137 a 247 c 170 g 203 t
ORIGIN
Query Match 25.6%; Score 378.8; DB 165; Length 757;
Best Local Similarity 74.6%; Pred. No. 3.2e-109;
Matches 519; Conservative 0; Mismatches 167; Indels 10; Gaps 3;
Qy 190 gcttgagatgaaccgcggccgagcttagctagtgtgtgaggttaagtcctcacaaagcga 249
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 689 GCCCGAGGAGGGGCTCGCGTCTAATTAGCTAGTGTGTGAGCAATAGCTTACCAAGCGGA 630
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 250 cgatcgtagccgcgctgagaggtgtccggtccacaaatgaactgagacacggtccatcac 309
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 629 TGATCAGTAGCTGTCGCGAGAGGATGATCAGCCACACTGGGACTGAGACACGCCCCAGAC 570
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 310 tctacggggggcagcaggttaagaatcttgcctcaatgggggggaaacccctgaagcagcagc 369
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 569 TCCTACGGGAGGAGCAGTAGTGGGAAATTTCCGCAATGGGCGCAAGCCTGACGAGCAATG 510
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 370 ccgctgaacgaagaagctcttcgattgttaaagttcattaggcagagaaataaagcagc 429
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 509 CCGCGTGGAGGTAGAGCCCTACGGGTCTGAACTCTTTTCCAG--AGAAGAAGCAAT 452
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 430 aatgtgatgtgtacctgacctaaagcaccggtctaactacgtgccagcaccggtgtaat 489
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 451 GACGGTATCTGGGA-----ATAAGCATCGGCTAACTCTGTGCCAGCAGCCGCGTAAAT 398
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 490 acgtatgtgcaacgctgtgttcggaatcattggcgctaaaagggtgctagcggatttgt 549
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 397 ACAGAGGATGCAAGCGTTATCCGGAATGATTGGCGGTAAACGCTCTGTAGTGGCTTTT 338
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 550 aatcaggtgtgaaactgcggctcaaccgtgctgcctgcacttgaaactacaaagctcg 609
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 337 AAGTCCCGCTCAATCCAGGGCTCAACCTTGACAGCGCGGTGGAACCTACCAAGCTTG 278
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 610 agtttggagaggaagtggaattcccggtgtgagcgtgaaatcgttagatatctgaggg 669
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 277 AGTACGGTAGGGGAGAGGGAATTTCCGGTGGAGCGGTGAATCGGTAGAGATCGGAAG 218
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 670 aacaccagtgcaagcagcacttgcctgcctcaaaactgacgctgagcgacgaagcgtgg 729
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 217 AACACCAACGGCGAAGACACTCTCTGGGCCGACACTGACACTGAGAGACGAAAGCTAGG 158
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 730 gtagtaacgggattagataccccggtaatccacgcccctaacgcttctaccagtttgt 789
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 157 GGAGCGAATGGATAGATATACCCAGTAGTCTTACCGCTAAACGATGATAGCGGCT 98
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 790 ggggttttaaccc--tcagtaacgaactaagcgtgattaaagtagcgcctgggactat 847
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 97 GTGCGTATCGACCCGTGCGAGTGTGTAGCTAACGCGTTAAAGTATCCCGCCCTGGGAGTAC 38
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 848 gctcgaagtgaaactcaaggaattgacgggggtgacggggg 883
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 37 GTTCCCAAGATGAACATCAAGAGGATTCACGGGGG 2
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 7
LOCUS B73824 676 bp DNA GSS 25-DEC-1997
DEFINITION T27K4TR TAMU Arabidopsis thaliana genomic clone T27K4, DNA
sequence.
ACCESSION B73824
VERSION B73824.1 GI:2712975
KEYWORDS GSS.
SOURCE
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.

```

REFERENCE 1 (bases 1 to 676)
 Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Linher,K., Golden,K.,
 Berry,K., Granger,D., Suh,E., Wible,C., Adams,M.D. and Venter
 J.C.
 TITLE A BAC End Sequence Database for Identifying Minimal Overlaps in
 Arabidopsis Genomic Sequencing. Update 3
 JOURNAL Unpublished (1997)
 COMMENT Other_GSSs: T27K4TF
 Contact: Steve Rounsley
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: rounsley@tigr.org
 Seq primer: M13 Reverse
 Class: BAC ends
 High quality sequence stop: 676.

FEATURES
 source
 1..676
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="T27K4"
 /clone_lib="TAMU"
 /sex="hermaphrodite"
 /note="vector: BelORACII; Site_1: HindIII; Site_2: HindIII
 ; Produced by Rod Wing"
 BASE COUNT 170 a 164 c 213 g 128 t 1 others
 ORIGIN

Query Match 25.5%; Score 376.6; DB 175; Length 676;
 Best Local Similarity 76.2%; Pred. No. 1.6e-108;
 Matches 515; Conservative 0; Mismatches 155; Indels 6; Gaps 4;

Qy 610 agtttgagagcaagtggaattccaggtgtagcgtggaatcgtagatctggagg 669
 Db 1 AGTACGGTAGGGGACAGAGGAATTTCCGGTGGAGCGGTAATCGGTAGAGATCGGAAG 60
 Qy 670 aacaccagtggcgaaggcgacttctggtctcaaaactgacgtgagcacaagcggtgg 729
 Db 61 AACACCAACGGCAAGACACTCTGCTGGCCGACACTGACACTGAGAGCAAGACTAGG 120
 Qy 730 gtagtaacgggattagatcccccggtgaatccacgcccctaaccctgtgtctaccagttgt 789
 Db 121 CGAGCGAATGGGATTAGATACCCCAAGTAGTCTTACCGCTAAACGATGGATACTAGCGCT 180
 Qy 790 gggggttttaaccc--tcagtaacgaacctaacggtattaaatagacgcgctgggaactat 847
 Db 181 GTGCGTATCGACCCGTGAGTGCCTAGCTAACCGGTTAAGTATCCCGCCCTGGGGAGTAC 240
 Qy 848 gctcgcaagagtgaactcaaggaattgacggggttcgcgcacaaagcggtggagcatgtg 907
 Db 241 GTTCGAAGATGAACACTCAAGGAATTGACGGGGCCCGCACAAAGCGGTGAGCATGTG 300
 Qy 908 gtttaattcgatatacccccaaacctcaacctgggttgatggaatcctgaate-atgtg 966
 Db 301 GTTTAAATTGATGCAAGCGAAGAACCTTTACAGGGCTTGACATGCCCGGAATCCTCTTG 360
 Qy 967 agagatatagccttcggg--cagattcacaggtgctgcatggttgcctcaagctcgt 1024
 Db 361 AAAGAGAGGGGTGCTTCGGGAACCGCGACACAGGTGGTGCATGGCTGCTCGCTCAGCTCGT 420
 Qy 1025 gtcgtgagatgttgggttaagtcccgcaacgagcgcaacccctatcgtatgttgcctacc- 1083
 Db 421 GCCGTAAAGGTGTTGGGTTAAGTCCGCAACGAGCGCAACCCCTCGTGTAGTTGCCACCG 480
 Qy 1084 ttaagttgggcacttggtagcaaaactgcggtgcacaaacccgaggaagcggtggatgcagt 1143
 Db 481 TTGAGTTTGGAAACCTGCAACAGACTGCCGGTGATAAAGCCGGGAAGGTGAGGATGACGT 540
 Qy 1144 caaatccctcatggcctttatgtccaggggccaacacacgctgtctacaatggccgatacagagg 1203

Db 541 CAAGTCATCATGCCCTTATGCCCTGNGCGACACACGTCGTACATGCCCGGACAAAAG 600
 Qy 1204 gtcgccaaactcgcaagggagcgtaactctctaaaagtcggtcccaagttcgcgattgggttc 1263
 Db 601 GTCGGATCCCGGAGGTGAGCTAACTCCAAAACCCGCTCTCAGTTCCGATTGCAGCG 660
 Qy 1264 tgcaactcgaccccat 1279
 Db 661 TGCAACTCGCTGCAT 676

RESULT 8
 LOCUS AQ956298/c 693 bp DNA GSS 28-JAN-2000
 DEFINITION LERAJ19TR LERA Arabidopsis thaliana genomic clone LERAJ19, DNA
 sequence.
 ACCESSION AQ956298
 VERSION AQ956298.1 GI:6783999
 KEYWORDS GSS.
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 Brassicales; Brassicaceae; Arabidopsids.
 REFERENCE 1 (bases 1 to 693)
 AUTHORS Buell,C.R., Lin,X., Pai,G., Barnstead,M., Bowman,C., Utterbach,T.,
 Feldblum,T., Liang,F., Creasy,T. and Fraser,C.M.
 TITLE Genomic survey sequencing of Landsberg erecta ecotype of
 Arabidopsis thaliana and identification of sequence-based
 polymorphisms
 JOURNAL Unpublished (2000)
 COMMENT Contact: Xiaoying Lin
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: atetigr.org
 For additional information, see <http://www.tigr.org/tdb/at/at.html>
 Similar to A. thaliana chloroplast sequence (GB:AP000423)
 Seq primer: TR
 Class: shotgun.

FEATURES
 source
 1..693
 /organism="Arabidopsis thaliana"
 /strain="Landsberg erecta"
 /db_xref="taxon:3702"
 /clone="LERAJ19"
 /clone_lib="LERA"
 /note="Organ: Leaf; Vector: phosI; Total genomic DNA was
 sheared to 0.9-1 kbp before ligation."
 BASE COUNT 136 a 221 c 157 g 178 t 1 others
 ORIGIN

Query Match 24.4%; Score 360.4; DB 165; Length 693;
 Best Local Similarity 77.2%; Pred. No. 2.2e-103;
 Matches 477; Conservative 0; Mismatches 136; Indels 5; Gaps 3;

Qy 453 aagcaccgctaactacatcgctccagcagcggttaatacgtatgctgaagcgtgttgcg 512
 Db 626 AAGCATCGCTAACTCTGTGCCAGCAGCGCGGTAAACAGAGGATGCNAGGTTATCGG 567
 Qy 513 gaatcattggcgtaaaaggggtgcgtaggcgtattgttaagtcaggtgtgaaactgcggg 572
 Db 566 GAATGATTGGCGTAAGCGTCTGTAGTGGCTTTTAAAGTCGCCCGTCATAATCCAGGG 507
 Qy 573 ctcaaacccgtggcctgcactgaaactacaactcgtggtggtttggagagggcaagtggaaat 632
 Db 506 CTCACCCCTGGACACAGCGGTGGAACCTACCAAGCTTGATACGTTAGGCGTAGGGCAGAGGAAT 447
 Qy 633 tccagggttaagcgtgaaatcgtagatatcttgagggaacaccacagtcggcgaagcgactt 652

Db 446 TTCCGGTGGAGCGGTGAATGCTAGAGATCGGAAGAACACCAACGCGGGAAGCACTCT 387
Qy 693 gctggtcctaaactgacgtgagcgacgaaagcgtggttagtaaacggtgattagataccc 752
Db 386 GCTGGGCGGACACTGACACTGACAGACGAAGAGCTAGGGGAGCGAATGGGATTAGATACCC 327
Qy 753 cggtaataccacgcccctaaacgctgtctaccagttgttgggggttttaacccc--tcagtaa 810
Db 326 CAGTAGTCTTAGCGGTAAACGATGGATAGTACCTAGCGCTGTCCGTATCGACCCGTGCGTGC 267
Qy 811 cgaactaaacggtattagtagacgctgaggaactatgctcgaagagtgaaactcaag 870
Db 266 TGTAGCTTAACGCGGTAAAGTATCCCGCTGGGAGTAGTACGTTTCGCAAGAATGAACCTCAAG 207
Qy 871 gaattgacgggggtccgcacaaagcgggtggagcatgtggttttaattcgtatgataccccc 930
Db 206 GAATGACGGGGGCCCCACAAAGCGGTGGAGCATGTGCTTAAATTCGATGCAAGCGAAG 147
Qy 931 aacctcactgggcttgacatggaatcgatc--atgtagagatatagaccttcgg-- 987
Db 146 AACCTTACAGGGCTTGACATGCGCGCAATCCCTCTTGAAGAGAGAGGGGTCCTTCGGGAA 87
Qy 988 cagattcacaggtgctcatgttctgctcagctcgtgctgagatgttgggttaagtc 1047
Db 86 CCGGACACAGGTGTCATGGCTGTCGTCAGCTCGTCCGCTAAGGTGTTGGGTAAAGTC 27
Qy 1048 ccgcaacgagcgcaacc 1065
Db 26 CCGCAACGAGCGCAACCC 9

RESULT 9
AQ969911/c 676 bp DNA GSS 28-JAN-2000
LOCUS LERJR28TF LERG Arabidopsis thaliana genomic clone LERJR28, DNA
DEFINITION sequence.
ACCESSION AQ969911
VERSION AQ969911.1 GI:6797612
KEYWORDS GSS.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 676)
AUTHORS Buell,C.R., Lin,X., Pai,G., Barnstead,M., Bowman,C., Utterbach,T.,
Fel'dblyum,T., Liang,F., Creasy,T. and Fraser,C.M.
TITLE Genomic survey sequencing of Landsberg erecta ecotype of
Arabidopsis thaliana and identification of sequence-based
polymorphisms
JOURNAL Unpublished (2000)
COMMENT Contact: Xiaoying Lin
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: atetigr.org
For additional information, see <http://www.tigr.org/tdb/at/at.html>
Similar to A. thaliana chloroplast sequence (GB:AF000423)
Seq primer: TF
Class: Shotgun.
Location/Qualifiers
1. 676
/organism="Arabidopsis thaliana"
/strain="Landsberg erecta"
/db_xref="taxon:3702"
/clone="LERJR28"
/clone_lib="LERG"
/note="Organ: Leaf; Vector: pUC19JK; Total genomic DNA was
sheared to 0.4-0.7 kbp before ligation."
BASE COUNT 133 a 207 c 177 g 159 t
ORIGIN

Query Match 24.3%; Score 359.6; DB 165; Length 676;
Best Local Similarity 76.7%; Pred. No. 4e-103;
Matches 517; Conservative 0; Mismatches 149; Indels 8; Gaps 6;

Qy 755 gtaattccacgcccctaaacgctgtctaccagttgttgggggttttaacccc--tcagtaa 812
Db 673 GTAGTCTTAGCGGTAAACGATGGATAGTACCTAGCGCTGTGCGTATCGACCCGTGCGTGTG 614
Qy 813 aacctaacggtattagtagacgctgaggaactatgctcgaagagtgaaactcaagga 872
Db 613 TAGTAAACGCGGTAAAGTATCCCGCTGGGAGTAGTACGTTTCGCAAGAATGAACCTCAAGGA 554
Qy 873 attgacgggggtccgcacaaagcgggtggagcatgtggttttaattcgtatgataccccc 932
Db 553 ATTGACGGGGGCCCCACAAAGCGGTGGAGCATGTGCTTAAATTCGATGCAAGCGAAGAA 494
Qy 933 cctcactcactgggcttgacatggaatcgatc--atgtagagatatagaccttcgg--ca 989
Db 493 CTTTACACAGGGCTTGACATGCGCGCAATCCCTCTTGAAGAGAGAGGGTGCCTTCGGAACG 434
Qy 990 gattcacaggtgctcatgttctgctcagctcgtgctgagatgttgggttaagtc 1049
Db 433 CCGGACACAGGTGTCATGGCTGTCGTCAGCTCGTCCGCTAAGGTGTTGGGTAAAGTC 374
Qy 1050 gcaacgagcgcaaccctcactatgctgttctgctcagctcgtgctgagatgttgggttaagtc 1108
Db 373 GCAACGAGCGCAACCCCTCGTGTGTTAGTTGCCACCGCTTGAGTTGGAACCCCTGAACAGAT 314
Qy 1109 gccggtgacaaacccgaggaagcggggtgagcgtcaaatcctcctatggcctttatgtcca 1168
Db 313 GCCGGTATTAAGCCGGAGAGGTGAGGATGACGTCAAGTCATCATGCTCCCTTATGCTCCT 254
Qy 1169 gggccacacacgtctacaaatggccgagatcacagaggtgcgaactcgaagaggagcta 1228
Db 253 GGGCGACACACGCTGTACAAATGGCGGGAACAAAGGTGCGATCCCGGAGGTGAGCTA 194
Qy 1229 atctctaaaagtcggtcccgagttcgagattggggtctgcaactcgaccctcgaagtcgga 1288
Db 193 ACTCCAAAACCCGCTCCTCAGTTCGGATTGCGAGGTGCAACTCGCTGTCATCAAGCCGGA 134
Qy 1289 atcgtagtaatacgcggagtag--catgccgcggtgtaatacgttcccgaccttgatacca 1347
Db 133 ATCGTAGTAAATCGCGGTGAGCATACGCGTGAATTCGTTCCCGGCGCTGTGATACCA 74
Qy 1348 cgcgcgctcacaccacactgaatgggagcagcccgagtggtcttcttaccgctaaaggag 1407
Db 73 CCGCCGCTCACACTATGGGAGCTGGCCATGCCCGAAGTCGT-TACCTTAACCGCAAGGAG 15
Qy 1408 acagactactaaag 1421
Db 14 GGGATGCCGAAGG 1

RESULT 10
AQ969996/c 659 bp DNA GSS 28-JAN-2000
LOCUS LERJR90TF LERG Arabidopsis thaliana genomic clone LERJR90, DNA
DEFINITION sequence.
ACCESSION AQ969996
VERSION AQ969996.1 GI:6797697
KEYWORDS GSS.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 659)
AUTHORS Buell,C.R., Lin,X., Pai,G., Barnstead,M., Bowman,C., Utterbach,T.,
Fel'dblyum,T., Liang,F., Creasy,T. and Fraser,C.M.
TITLE Genomic survey sequencing of Landsberg erecta ecotype of
Arabidopsis thaliana and identification of sequence-based
polymorphisms

JOURNAL Unpublished (2000)
 COMMENT Contact: Xiaoying Lin
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: at@tigr.org
 For additional information, see <http://www.tigr.org/tdb/at/at.html>
 Similar to A. thaliana chloroplast sequence (GB:AP000423)
 Seq primer: TF
 Class: shotgun.

FEATURES
 source Location/Qualifiers
 1..659
 /organism="Arabidopsis thaliana"
 /strain="Landsberg erecta"
 /db_xref="taxon:3702"
 /clone="LERJ90"
 /clone_lib="LERG"
 /note="Organ: Leaf; vector: pUC19JK; Total genomic DNA was sheared to 0.4-0.7 Kbp before ligation."
 130 a 206 c 163 g 160 t

BASE COUNT 130 a 206 c 163 g 160 t

ORIGIN

Query Match 24.1%; Score 356.6; DB 165; Length 659;
 Best Local Similarity 77.3%; Pred. No. 3.6e-102;
 Matches 497; Conservative 0; Mismatches 139; Indels 7; Gaps 5;

Qy 692 tgcgtgctcaaaactgacgtgagcgacgaagcggtgtagtaaacgggattagatacc 751
 Db 643 TGCCTGGCCGACATGACATGACAGACGAAGGTAGGGGAGCAATGGGATTTAGATACC 584
 Qy 752 ccggtaatccacccctaaacgttgtctaccagttgttgggggttttaacc--tcagta 809
 Db 583 CCAGTAGTCTAGCGTAAACGATGATACCTAGGCGGTGTCGCTATCGACCCGTGCAGTG 524
 Qy 810 acgaacctaaacgattagtagacgcctggtggtgactatgctcgcaagagtgaacctcaaa 869
 Db 523 CTGTAGTAAACGGTTAAGTATCCCGCTGGGGAGTAGCTTCGGAAGAAATGAATCAAA 464
 Qy 870 ggaattgacgggggtccgcacaaacgggtgagcagtgtgtttaaattcgtatgatacccaaa 929
 Db 463 GGAATTGACGGGGCCCGCACACGGGTGGAGCATGTGTTTAAATTCGATCAAGCGAA 404
 Qy 930 aaacctcacctgggttgacatgatctgaatc-atgtagagatatatgagccttcggg- 987
 Db 403 GAACCTTACCAGGGCTTGACATCGCGCAATCCTCTTGAAGAGAGGGGTGCTTCGGGA 344
 Qy 988 -cagattcacaggtgctgcatgtgtgctcagctcgtgtcgtgagatgttggttaagt 1046
 Db 343 ACGGGGACACAGGTGGTGCATGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 284
 Qy 1047 cccgcaacgagcgcaacccctctgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1105
 Db 283 CCCGCAACGAGCGCAACCTCTGTTAGTTGCGCACCGCTTGAGTTTGAACCTTGACAG 224
 Qy 1106 actgcccgttgacaaacccgggaggaaggcggtgacgtcgaatcctcctgcttattgt 1165
 Db 223 ACTGCCGTGTAAAGCCGGAGGAGGTGAGGATGACGTGCAAGTATCATGCTGCTTATGC 164
 Qy 1166 cgaaggccacacgctgctacaaatggccgatacagaggggtgcgcaactcgcaagaggag 1225
 Db 163 CTTGGGGACACAGCTGCTACAAATGGCCGGGACAAAGGGTCCGATCCCGGAGGGTGA 104
 Qy 1226 ctaactctctaaagtcgggtccaggttcggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1285
 Db 103 CTAACCTCAAAACCCGCTCCTGATTTGCGATTTGAGGCTGCACTAGCCTGCATGAAGCC 44
 Qy 1286 ggaatcgtgtagtaatcgcggatcac-gatgccgcggtggaatc 1327
 Db 43 GGAATCGCTAGTAATCGCGGTCAGCATACGCGGGTGAATTC 1

RESULT 11
 AW221851 677 bp mRNA EST 07-DEC-1999
 LOCUS EST298662 tomato fruit red ripe, TAMU Lycopersicon esculentum cDNA
 DEFINITION clone cLEN4113, mRNA sequence.
 ACCESSION AW221851 GI:5533535
 VERSION AW221851.1
 KEYWORDS EST.
 SOURCE tomato.
 ORGANISM Lycopersicon esculentum
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids
 I; Solanales; Solanaceae; Solanum; Lycopersicon.
 REFERENCE 1 (bases 1 to 677)
 AUTHORS Alcala,J., Vrebalov,J., White,R., Matern,A.L., Holt,I.E., Liang,F.,
 Upton,J., Hansen,T., Craven,M.B., Bowman,C.L., Ahn,S., Ronning,C.M.,
 Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
 TITLE Generation of ESTs from tomato fruit tissue
 JOURNAL Unpublished (1999)
 COMMENT Contact: David Frisch
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 4366
 Fax: 864 656 4293
 Email: dfrisch@CLEMSON.EDU
 5 prime sequence.

FEATURES
 source Location/Qualifiers
 1..677
 /organism="Lycopersicon esculentum"
 /cultivar="TA496"
 /db_xref="taxon:4081"
 /clone="cLEN4113"
 /clone_lib="tomato fruit red ripe, TAMU"
 /tissue_type="pericarp"
 /dev_stage="red ripe (7-20 days post-breaker)"
 /note="Vector: p Bluescript SK(-); Site 1: EcoRI; Site 2:
 XhoI; supplier: Giovannoni; Fruit were tagged at the
 breaker stage (first sign of lycopene accumulation on the
 blossom end of the fruit) and harvested 7 days
 post-breaker (fully red-ripe), 10 days post breaker, and
 20 days post-breaker (over-ripe). 20 day fruit which
 showed external or internal signs of pathogenesis were
 discarded. Fruit were cut in half and the seeds and
 locules were discarded prior to freezing the pericarp."
 156 a 178 c 213 g 130 t

BASE COUNT 156 a 178 c 213 g 130 t

ORIGIN

Query Match 23.9%; Score 352.4; DB 87; Length 677;
 Best Local Similarity 77.0%; Pred. No. 7.8e-101;
 Matches 493; Conservative 0; Mismatches 141; Indels 6; Gaps 5;

Qy 834 cgcctgggaactatgctcgcaagagtgaactcaagaattgacgggggtcgcaaga 893
 Db 1 GCCTTGGGGAGTACGCTGCGAAGACCGAAATCAAGGAATTCACGGGGGCTGCACAA 60
 Qy 894 cggtagagcatgtgttttaattcgtatgatacccaaaacctcacctgggttgatagg 953
 Db 61 CGGTGGACATGTGGTTTAAATTCGATACACCGCGCAAAACCTTACCAGGGCTTGACATGC 120
 Qy 954 atctgaatc-atgtagagatatatgaccttcggg--cagattcacaggtgctgcatggt 1010
 Db 121 CGCGAAATCCTCTTGAAGAGAGGGGTGCTTCGGGAACCGGACACAGGTGTGTCATGTC 180
 Qy 1011 tgcgtcagctcgctgctgagatgtttgggttaagtccgcaacgagcgcaaccctatc 1070
 Db 181 TGTCGTACAGTCGTCGCGTAAAGGTGTTGGGTAGTCCCGCAACGACGCAACCCCTCTG 240
 Qy 1071 gtagttgtcta-ccttaagtggcactggttacgaaactgcccgtgacaaacccgagagaa 1129
 Db 241 TTTAGTTCCCATCGTTGAGTTTGAACCTTGCAGACTGCGGGCGCATAGCCGGAGGAA 300

```
QY 1130 ggcggggaagtcgcaaatcctcctatggcctttatgtctccagggccacacacgctgtacaaat 1189
Db 301 GGTGAGTACAGTCAAGTCAATCATCCCTTATGCCCTGGCGACACACGCTCTACAAT 360
QY 1190 ggcgcatacagaggtcgccaactcgcaagaggagactaatctctaaagtgcgtcccgag 1249
Db 361 GGCCTGGGCAAGAGGTGCGGATCCCGAGGGTGAGCTAACCCCAAAACCCGCTCCTCAG 420
QY 1250 ttgcgattgggtctgcaactcgaccctcgaagtgcggaatcgctagtaatcgcgatca 1309
Db 421 TTTCGATTGACGCTGCAACTCGCTTCGATGAAGCGCGAATCGCTAGTAATCGCCGTCA 480
QY 1310 g-catgccgcgggtgaatacagttccccggacctgtacacaccgcccgtcacacacacctgag 1368
Db 481 GCCATACGGCGGTGAATTCGTTCCCGGGCTTGTACACACCGCCGCTCACACTATGGGAG 540
QY 1369 tggggagaccgcgaagtgtctttgtaaccgtgaaggaagacagactactaagtgaaact 1428
Db 541 CTGGCATGCCCGAAGTCGT-TACCTTACCGCAAGAGGGGGATGCCGANGCAGGGCT 599
QY 1429 cgtaaagggtgaagtcgtaaacgaagtcacggttaaactcga 1468
Db 600 AGTGACTGGAGTCAAGTCGTAAACAGGTAGCCGCTACTGGA 639

RESULT 12
AG019017/c 616 bp DNA GSS 10-OCT-1999
LOCUS Homo sapiens genomic DNA, 21q region, clone: B753B2B067F, genomic
DEFINITION survey sequence.
ACCESSION AG019017
VERSION AG019017.1 GI:6017503
KEYWORDS GSS.
SOURCE Homo sapiens DNA, clone: B753B2B067F.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 616)
Hattori,M., Fujiyama,A., Ishii,K., Toyoda,A., Taylor,T.,
Park,H.-S., Yada,T., Watanabe,H. and Sakaki,Y.
Homo sapiens genomic DNA, chromosome 21q
Published Only in DataBase (1999) In press
2 (bases 1 to 616)
Hattori,M., Fujiyama,A., Ishii,K., Toyoda,A., Taylor,T.,
Park,H.-S., Yada,T., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (10-OCT-1999) to the DDBJ/EMBL/GenBank databases.
Masahira Hattori, RIKEN Genomic Sciences Center(GSC) c/o Kitasato
University, 1-15-1 Kitasato, Sagamihara 228-8555, Japan
(E-mail:hattori@hgc.ims.u-tokyo.ac.jp, Tel:042-778-9923,
Fax:042-778-9924)
FEATURES
Location/Qualifiers
source
1. .616
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosomes="21"
/clone="B753B2B067F"
/map="21q"
BASE COUNT 131 a 187 c 145 g 153 t
ORIGIN

Query Match 23.7%; Score 350.6; DB 190; Length 616;
Best Local Similarity 78.9%; Pred. No. 2.8e-100;
Matches 468; Conservative 0; Mismatches 119; Indels 6; Gaps 4;

QY 872 aattgacggggccgcacaaagggtgagacatggttttaattcgaatgataccccaataa 931
Db 616 AATTGACGGGGCCGCCACAAAGCGGTGACATGTGTTTAATTTCGATCAACCCGAAGA 557
QY 932 acctcactgggcttgacatggatctgtaaatcat-gtagagatatagagctctggg--c 988
Db 556 ACCTTACTTGCTCTTGACATCCACGGGAAGTTTTCAGAGATGAGAAATGTGCTTCGGGAAC 497
```

```
QY 989 agattcacagggtgctgcatgggttgcgtcagctcgctgctgagatgttggttaagtcc 1048
Db 496 CGTGACACAGGTGCTGCATGGGTGCTCGTCAGCTCGGTGTTGTAATGTTGGGTAAAGTCC 437
QY 1049 cgcacaacgacgaacccctctatcgatgttgctacc--ttaagttgggcaactggtagcaaa 1106
Db 436 CGCAGGAGCGCAACCTTATCTTTTGGCCAGCGGTCCGCCCGGGAACCTCAAGAGGAGA 377
QY 1107 ctgcgggtgacaaacgggagggcgggatgacgtcaaatctcatggtccctttatgtc 1166
Db 376 CTGCCAGTGATAAATGAGGAGGTGGGNTGACGTCAAGTCATCATGTCCTTACGAC 317
QY 1167 caggggccacacgctgtacaaatgcccgtacagaggggtcgccaaactcgcaagagggagc 1226
Db 316 CAGGGGTACACACGCTCTACAATGGCGCATACAAAGAGAAGCGACCTCGCGAGCAAGC 257
QY 1227 taatctctaaagtcgctcccgatctcgagttgggtgctcgaactgaccccatgaagtccg 1286
Db 256 GGACCTCATAAAGTCGTCTAGTCCGGATTGGAGTCTGCAACTCGACTCCATGAAGTCG 197
QY 1287 gaatcgtagtaatcgatcgatcagcatcgccggtgaatcagcttcccgacctgttacac 1346
Db 196 GAATCGCTAGTAATCGTGTGATGCAATGCCACGTGAATACGTTCCCGGCTTGTACAC 137
QY 1347 accgccgtcacaccctgagtggtgggagaccgccgagtggtcttcttgaacgtagga 1406
Db 136 ACCGCCGTCACACCATGGGAGTGGGTTCGAAAGAAAGTAG-GTAGCTTAACCTTCGGGA 78
QY 1407 gacagactactaagggtgaactcgtaaaagggtgaagtcgtaacaaaggtacc 1459
Db 77 GGGCGCTTACCACCTTTGTGATTTCATGACTGGGTGAAGTCGTAAACAAGGTAAC 25

RESULT 13
AG056297 675 bp DNA GSS 28-JAN-2000
LOCUS LERAJ19F LERA Arabidopsis thaliana genomic clone LERAJ19, DNA
DEFINITION sequence.
ACCESSION AG056297
VERSION AG056297.1 GI:6783998
KEYWORDS GSS.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
REFERENCE Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
AUTHORS Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 675)
Buell,C.R., Lin,X., Pai,G., Barnstead,M., Bowman,C., Utterbach,T.,
Feldblyum,T., Liang,F., Creasy,T. and Fraser,C.M.
Genomic survey sequencing of Landsberg erecta ecotype of
Arabidopsis thaliana and identification of sequence-based
polymorphisms
Unpublished (2000)
Contact: Xiaoying Lin
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: at@tigr.org
For additional information, see http://www.tigr.org/tldb/at/at.html
Similar to A. thaliana chloroplast sequence (GB:AP000423)
Seq primer: TF
Class: shotgun.
Location/Qualifiers
source
1. .675
/organism="Arabidopsis thaliana"
/strain="Landsberg erecta"
/db_xref="taxon:3702"
/clone="LERAJ19"
/clone_lib="LERA"
/notes="Organ: Leaf; Vector: phOS1; Total genomic DNA was
sheared to 0.9-1 Kbp before ligation."
```


JOURNAL: Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE: 20202663
COMMENT: Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel.: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st3-QV0-TT0009-191
199-056-b02&t3=1999-11-19&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 5
High quality sequence stop: 659.

FEATURES
Location/Qualifiers
I..659
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="TT0009"
/dev_stage="Adult"
/note="Organ: testis; Vector: puc18; Site_1: Sma1; Site_2:
Sma1; A mini-library was made by cloning products derived
from OKRESTES PCR (U.S. Letters Patent application No. 196
716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."
BASE COUNT 152 a 146 c 234 g 126 t 1 others
ORIGIN

Query Match 23.3%; Score 344.; DB 95; Length 659;
Best Local Similarity 74.4%; Pred. No. 3.6e-98;
Matches :490; Conservative 0; Mismatches 156; Indels 13; Gaps 4;

QY 254 ccgtagccggcctgagagggtgccggccacaatgaactgagacagcgctccactcct 313
| | | | |
Db 13 CGGTGCTCGTGTGAGAGGATGATCAG-CACACTGGGACTGAGACACGCCGCACTCCT 71

QY 314 acgggaggcagcttaagaattcttgcataatgggggaaacctgaagcagcgcgcgc 373
| | | | |
Db 72 ACGGGAGGCAGCAGTGCGGGAATTGGACAATGGGGCAACCCCTGATCCAGCAATGCGCG 131

QY 374 gtgaacagaagaaggttcgcgaatttaagttaagttcatlaagcaggagaaaaataagcagaatg 433
| | | | |
Db 132 GTGAGTGAAGAAGGGCTTAGGGTTTGTAAAGCTCTTTTCGCCCTGACGATGATACGGTAA 191

QY 434 tgatgatgttacgtgctaagaaccgcccgaactactcgtgcacagcagcgcgcgttaacgt 493
| | | | |
Db 192 CCGGAGAAG-----AAGCCCCGCTTAATCCGTGCCAGCAGCCGCTAATACGG 241

QY 494 atggtgcaagcgttgttgogaatcattggcgtaaagggtcgtaggcggaatttgttaagt 553
| | | | |
Db 242 AAGGGCTAGCGTTGTTTCGATTACTGGCGTAAAGGCATGTAGGCGCGGATCAAGT 301

QY 554 cagggtgaaaactcggggctcaaccctgccccctcacctgaaactcaaacatcgtggagtt 613
| | | | |
Db 302 GGGCGGTGAAGAGCCCTGGGCTCAACCTGGGAACCTCCGCTTCAGACTGGTTTCGTCGAG-T 360

QY 614 tgggagaggcaagtgaattccagggttagcgtgaaatgcgtagatatcctggaggaaaca 673
| | | | |
Db 361 TCGGAGAGGGTGTGGGAATCCCAGTGTAGAGTGAATTCGTAGATATTGGGAGGACACA 420

QY 674 ccagttggcgaagcgcaactgtcggtctcaaaactgacgcgtgagcagcaaacgctgggttag 733
| | | | |
Db 421 CCGGTGCCAAGCGCGCCACCCTGGCCCGAGAGACTGACGCTCAAGTCGGAACCGCTGGGAG 480

QY 734 taacaggagattagatacccccgtgaatccacgcctcaaacgttgtctaccagttgttggtggg 793
| | | | |
Db 481 CAACACAGGATTAGATACCTGGTGTAGTCCACGCTGTAAACGATGGGTGTAGTCTGCTCGGG 540

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel.: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st3-QV0-TT0009-191
199-056-b02&t3=1999-11-19&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 5
High quality sequence stop: 659.

FEATURES
Location/Qualifiers
I..659
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="TT0009"
/dev_stage="Adult"
/note="Organ: testis; Vector: puc18; Site_1: Sma1; Site_2:
Sma1; A mini-library was made by cloning products derived
from OKRESTES PCR (U.S. Letters Patent application No. 196
716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

BASE COUNT 152 a 146 c 234 g 126 t 1 others
ORIGIN

Query Match 23.3%; Score 344.; DB 95; Length 659;
Best Local Similarity 74.4%; Pred. No. 3.6e-98;
Matches :490; Conservative 0; Mismatches 156; Indels 13; Gaps 4;

QY 254 ccgtagccggcctgagagggtgccggccacaatgaactgagacagcgctccactcct 313
| | | | |
Db 13 CGGTGCTCGTGTGAGAGGATGATCAG-CACACTGGGACTGAGACACGCCGCACTCCT 71

QY 314 acgggaggcagcttaagaattcttgcataatgggggaaacctgaagcagcgcgcgc 373
| | | | |
Db 72 ACGGGAGGCAGCAGTGCGGGAATTGGACAATGGGGCAACCCCTGATCCAGCAATGCGCG 131

QY 374 gtgaacagaagaaggttcgcgaatttaagttaagttcatlaagcaggagaaaaataagcagaatg 433
| | | | |
Db 132 GTGAGTGAAGAAGGGCTTAGGGTTTGTAAAGCTCTTTTCGCCCTGACGATGATACGGTAA 191

QY 434 tgatgatgttacgtgctaagaaccgcccgaactactcgtgcacagcagcgcgcgttaacgt 493
| | | | |
Db 192 CCGGAGAAG-----AAGCCCCGCTTAATCCGTGCCAGCAGCCGCTAATACGG 241

QY 494 atggtgcaagcgttgttgogaatcattggcgtaaagggtcgtaggcggaatttgttaagt 553
| | | | |
Db 242 AAGGGCTAGCGTTGTTTCGATTACTGGCGTAAAGGCATGTAGGCGCGGATCAAGT 301

QY 554 cagggtgaaaactcggggctcaaccctgccccctcacctgaaactcaaacatcgtggagtt 613
| | | | |
Db 302 GGGCGGTGAAGAGCCCTGGGCTCAACCTGGGAACCTCCGCTTCAGACTGGTTTCGTCGAG-T 360

QY 614 tgggagaggcaagtgaattccagggttagcgtgaaatgcgtagatatcctggaggaaaca 673
| | | | |
Db 361 TCGGAGAGGGTGTGGGAATCCCAGTGTAGAGTGAATTCGTAGATATTGGGAGGACACA 420

QY 674 ccagttggcgaagcgcaactgtcggtctcaaaactgacgcgtgagcagcaaacgctgggttag 733
| | | | |
Db 421 CCGGTGCCAAGCGCGCCACCCTGGCCCGAGAGACTGACGCTCAAGTCGGAACCGCTGGGAG 480

QY 734 taacaggagattagatacccccgtgaatccacgcctcaaacgttgtctaccagttgttggtggg 793
| | | | |
Db 481 CAACACAGGATTAGATACCTGGTGTAGTCCACGCTGTAAACGATGGGTGTAGTCTGCTCGGG 540

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 27, 2001, 08:22:56 ; Search time 207.51 Seconds
(without alignments)
39,827 Million cell updates/sec

Title: US-09-380-826A-2
Perfect score: 22
Sequence: 1 ttttgatcacaaagtgtgata 22

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_36:*

1:	/cgn2_2/gcgdata/geneseq/geneseq/NA1980.DAT:*
2:	/cgn2_2/gcgdata/geneseq/geneseq/NA1981.DAT:*
3:	/cgn2_2/gcgdata/geneseq/geneseq/NA1982.DAT:*
4:	/cgn2_2/gcgdata/geneseq/geneseq/NA1983.DAT:*
5:	/cgn2_2/gcgdata/geneseq/geneseq/NA1984.DAT:*
6:	/cgn2_2/gcgdata/geneseq/geneseq/NA1985.DAT:*
7:	/cgn2_2/gcgdata/geneseq/geneseq/NA1986.DAT:*
8:	/cgn2_2/gcgdata/geneseq/geneseq/NA1987.DAT:*
9:	/cgn2_2/gcgdata/geneseq/geneseq/NA1988.DAT:*
10:	/cgn2_2/gcgdata/geneseq/geneseq/NA1989.DAT:*
11:	/cgn2_2/gcgdata/geneseq/geneseq/NA1990.DAT:*
12:	/cgn2_2/gcgdata/geneseq/geneseq/NA1991.DAT:*
13:	/cgn2_2/gcgdata/geneseq/geneseq/NA1992.DAT:*
14:	/cgn2_2/gcgdata/geneseq/geneseq/NA1993.DAT:*
15:	/cgn2_2/gcgdata/geneseq/geneseq/NA1994.DAT:*
16:	/cgn2_2/gcgdata/geneseq/geneseq/NA1995.DAT:*
17:	/cgn2_2/gcgdata/geneseq/geneseq/NA1996.DAT:*
18:	/cgn2_2/gcgdata/geneseq/geneseq/NA1997.DAT:*
19:	/cgn2_2/gcgdata/geneseq/geneseq/NA1998.DAT:*
20:	/cgn2_2/gcgdata/geneseq/geneseq/NA1999.DAT:*
21:	/cgn2_2/gcgdata/geneseq/geneseq/NA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	22	100.0	22	19	V58901 Leptospiira rRNA ge
2	22	100.0	22	19	V58897 L. fainei nucleoti
3	22	100.0	1477	19	V58896 L. fainei nucleoti
c 4	16.8	76.4	2582	18	T90543 Maize male fertili
c 5	16.2	73.6	951	20	V84598 Human secreted pro
c 6	16.2	73.6	1291	19	V59636 Human secreted pro
c 7	16.2	73.6	1473	20	X61757 B. burgdorferi ant
8	16.2	73.6	2000	17	T17111 Acetyl-CoA-carboxy
9	16.2	73.6	2001	17	T39904 Maize acetyl CoA c
10	16.2	73.6	2001	17	T49816 EcoRI fragment of
11	16.2	73.6	3822	18	T64683 M. leprae gyrA pre
12	16.2	73.6	4346	14	Q42933 A3 maize ACCase cd

13	16.2	73.6	7470	17	T39905
14	16.2	73.6	7470	19	V29317
15	16.2	73.6	7470	21	Z49820
c 16	16.2	73.6	10811	19	V59091
17	15.8	71.8	843	20	Z15467
18	15.8	71.8	961	19	X14095
19	15.8	71.8	4146	21	Z51556
c 20	15.6	70.9	297	20	V88753
c 21	15.6	70.9	795	20	V72025
c 22	15.6	70.9	834	20	V72026
c 23	15.6	70.9	1075	18	T72715
c 24	15.6	70.9	1096	13	Q20498
c 25	15.6	70.9	1194	20	V72027
c 26	15.6	70.9	1560	18	T49864
c 27	15.6	70.9	1584	19	V28845
c 28	15.6	70.9	1761	21	Z45339
c 29	15.6	70.9	1989	20	V81481
c 30	15.6	70.9	2354	19	V50429
c 31	15.6	70.9	2651	20	X13369
c 32	15.6	70.9	7745	18	T79859
c 33	15.6	70.9	7745	18	T59308
c 34	15.6	70.9	7745	20	X04183
c 35	15.6	70.9	9824	14	Q40763
c 36	15.2	69.1	349	13	Q34809
c 37	15.2	69.1	493	20	X13759
c 38	15.2	69.1	1028	17	T36923
c 39	15.2	69.1	1848	16	Q86996
c 40	15.2	69.1	1848	16	Q86998
c 41	15.2	69.1	2202	17	T36922
c 42	15.2	69.1	2361	20	X04316
c 43	15.2	69.1	2598	18	T72286
c 44	15.2	69.1	2598	21	Z98878
c 45	15.2	69.1	5420	12	Q11643

ALIGNMENTS

RESULT 1
V58901
ID V58901 standard; DNA; 22 BP.
XX
AC V58901;
XX
DT 20-JAN-1999 (first entry)
DE Leptospiira rRNA gene nucleotide sequence.
XX
KW Infection; pathogenic Leptospiira; protective immunity; therapy;
diagnosis; SS.
XX
OS Leptospiira sp.
XX
PN WO9840099-A1.
XX
PD 17-SEP-1998.
XX
PF 06-MAR-1998; 98WO-AU00145.
XX
PR 07-MAR-1997; 97AU-0005494.
XX
PA (AGRI-) AGRIC VICTORIA SERVICES PTY LTD.
PA (PIGR-) PIG RES & DEV CORP.
XX
PI Chappel RJ;
XX
DR WPI; 1998-520791/44.
XX
PT New isolated pathogenic Leptospiira bacterium - useful for, e.g
developing products for conferring protective immunity, and for
prophylactic or therapeutic treatment
XX
PS Claim 15; Page 72; 94pp; English.

Maize acetyl CoA c
Maize ACCase enzym
Maize acetyl CoA c
Potato pollen cell
Human gene express
H. pylori GHPO 131
Human hypoxia resp
EST clone HK650.
Adenovirus PACTSG2
Adenovirus SCAR.RG
C. elegans inhibit
Encodes fibrinogen
Adenovirus PACSG2S
Callosellasma rhod
Human coxsackievir
DNA encoding a Bac
Black Creek Canal
Human coxsackievir
Enterococcus faeca
Genomic DNA encodi
SH2-mRev6 gene (s
Shrunken-2 gene va
Wild-type shrunken
Human type IV coll
Enterococcus faeca
Human OVCA2 tumour
Aminopeptidase O12
Aminopeptidase O12
Human OVCA1 tumour
Human secreted pro
Breast cancer mamm
MMTV env protein n
Partial human comp

XX This sequence represents a Leptospira DNA sequence isolated from the
 CC pathogenic Leptospira (LS) bacterium of the invention. The bacterium
 CC belongs to serogroup Hurstbridge or serovar hurstbridge or the species
 CC L. fainei. The LS bacteria can be used for conferring protective
 CC immunity against pathogenic LS bacteria in humans or animals. The
 CC bacteria can also be used for prophylactic or therapeutic treatment of LS
 CC infections. The DNAs and antibodies may also be used for detection and
 CC diagnosis of past or present LS infection.
 XX
 SQ Sequence 22 BP; 7 A; 2 C; 5 G; 8 T; 0 other;

Query Match 100.0%; Score 22; DB 19; Length 22;
 Best Local Similarity 100.0%; Pred. No. 0.047;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgttgatcacaaagatttgata 22
 Db 1 tgttgatcacaaagatttgata 22

RESULT 2
 V58897
 ID V58897 standard; DNA; 22 BP.

XX AC V58897;
 XX DT 20-JAN-1999 (first entry)
 XX DE L. fainei nucleotide sequence.

XX Infection; pathogenic Leptospira; protective immunity; therapy;
 KW diagnosis; ss.

XX Leptospira fainei.

XX WO9840099-A1.
 XX 17-SEP-1998.

XX 06-MAR-1998; 98WO-AU00145.
 XX 07-MAR-1997; 97AU-0005494.
 XX (AGRI-) AGRIC VICTORIA SERVICES PTY LTD.
 PA (PIGR-) PIG RES & DEV CORP.

XX Chappel RJ;

XX WPI; 1998-520791/44.

XX New isolated pathogenic Leptospira bacterium - useful for, e.g
 PT developing products for conferring protective immunity, and for
 PT prophylactic or therapeutic treatment

XX Claim 15; Page 70; 94pp; English.

XX This sequence represents a Leptospira DNA sequence isolated from the
 CC pathogenic Leptospira (LS) bacterium of the invention. The bacterium
 CC belongs to serogroup Hurstbridge or serovar hurstbridge or the species
 CC L. fainei. The LS bacteria can be used for conferring protective
 CC immunity against pathogenic LS bacteria in humans or animals. The
 CC bacteria can also be used for prophylactic or therapeutic treatment of LS
 CC infections. The DNAs and antibodies may also be used for detection and
 CC diagnosis of past or present LS infection.

XX Sequence 22 BP; 7 A; 2 C; 5 G; 8 T; 0 other;

Query Match 100.0%; Score 22; DB 19; Length 22;
 Best Local Similarity 100.0%; Pred. No. 0.047;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgttgatcacaaagatttgata 22
 Db 1 tgttgatcacaaagatttgata 22

RESULT 3
 V58896
 ID V58896 standard; DNA; 1477 BP.

XX AC V58896;

XX DT 20-JAN-1999 (first entry)

XX DE L. fainei nucleotide sequence.

XX Infection; pathogenic Leptospira; protective immunity; therapy;
 KW diagnosis; ss.

XX Leptospira fainei.

XX WO9840099-A1.

XX 17-SEP-1998.

XX 06-MAR-1998; 98WO-AU00145.

XX 07-MAR-1997; 97AU-0005494.

XX (AGRI-) AGRIC VICTORIA SERVICES PTY LTD.

XX (PIGR-) PIG RES & DEV CORP.

XX Chappel RJ;

XX WPI; 1998-520791/44.

XX New isolated pathogenic Leptospira bacterium - useful for, e.g
 PT developing products for conferring protective immunity, and for
 PT prophylactic or therapeutic treatment

XX Claim 15; Page 69-70; 94pp; English.

XX This sequence represents a Leptospira DNA sequence isolated from the
 CC pathogenic Leptospira (LS) bacterium of the invention. The bacterium
 CC belongs to serogroup Hurstbridge or serovar hurstbridge or the species
 CC L. fainei. The LS bacteria can be used for conferring protective
 CC immunity against pathogenic LS bacteria in humans or animals. The
 CC bacteria can also be used for prophylactic or therapeutic treatment of LS
 CC infections. The DNAs and antibodies may also be used for detection and
 CC diagnosis of past or present LS infection.

XX Sequence 1477 BP; 390 A; 334 C; 439 G; 314 T; 0 other;

Query Match 100.0%; Score 22; DB 19; Length 1477;
 Best Local Similarity 100.0%; Pred. No. 0.072;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgttgatcacaaagatttgata 22
 Db 154 tgttgatcacaaagatttgata 175

RESULT 4
 T90543/c
 ID T90543 standard; DNA; 2582 BP.

XX AC T90543;

XX DT 13-FEB-1998 (first entry)

XX DE Maize male fertility gene 233 Zm41-A.

KW Zm41-A; male fertility; hybrid seed; transgenic plant; antisense;
 KW ribozyme; male sterile; maize; Ms41-A; ds.

OS Zea mays.

XX Key Location/Qualifiers

XX exon 823..915

FT intron /*tag= a

FT exon 916..935

FT exon /*tag= b

FT exon 936..1090

FT intron /*tag= c

FT intron 1091..1167

FT exon /*tag= d

FT exon 1168..1332

FT exon /*tag= e

XX WO9723618-A1.

PN 03-JUL-1997.

XX 20-DEC-1996; 96WO-GB03191.

XX 21-DEC-1995; 95GB-0026218.

XX (GENE-) GENE SHEARS PTY LTD.

XX Baudot G, Garcia D, Hodge R, Perez P;

XX WPI; 1997-351055/32.

XX Nucleic acid encoding proteins involved in male fertility in plants

PT - used to control fertility and for production of hybrid seed

XX Claim 3; Fig 15; 85pp; English.

XX This DNA sequence comprises the maize Z33 Zm41-A gene. This is an

CC orthologue of Arabidopsis Ms41-A (see T90522), a gene that confers

CC male fertility. It was obtained following a database search for

CC sequences that showed homology to Ms41-A DNA, isolation of a

CC partial cDNA clone, and use of this clone to isolate Zm41-A genes

CC Z31 (T90542), Z33 (T90543) and Z35 (T90544) from genomic lambda

CC libraries. The genes show a high level of conservation. Z35 may

CC be derived from Z31 via genetic rearrangements, deletions and/or

CC insertions. Z33 has subsequent deletions from Z35 and is

CC truncated, having only exons 3, 5 and 6. The Ms41-A and Zm41-A

CC genes, antisense or ribozyme sequences can be used to produce

CC transgenic plants with controlled male fertility. Male sterile

CC plants are useful for hybrid seed production, particularly in

CC plants where restoration of fertility is not needed, e.g.

CC Brassicaceae, lettuce, spinach and onions.

XX SQ Sequence 2582 BP; 670 A; 512 C; 537 G; 863 T; 0 other;

Query Match 76.48; Score 16.8; DB 18; Length 2582;

Best Local Similarity 90.08; Pred. No. 26;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 tgttgatcaccaagatttga 20

||||| ||||||||| |||

Db 268 TGTTCATCACAGATCTGA 249

RESULT 5

V84598/c

ID V84598 standard; DNA; 951 BP.

XX V84598;

AC V84598;

XX 01-MAR-1999 (first entry)

DT Human secreted protein gene 188 clone HHPSP70.

XX

DE

XX Human; secreted protein; fusion protein; gene therapy; protein therapy;
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
 KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
 KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
 KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
 XX Homo sapiens.
 XX WO9854963-A2.
 PN 10-DEC-1998.
 PD 04-JUN-1998; 98WO-US11422.
 XX 18-DEC-1997; 97US-0070923.
 XX 06-JUN-1997; 97US-0048877.
 XX 06-JUN-1997; 97US-0048881.
 XX 06-JUN-1997; 97US-0048884.
 XX 06-JUN-1997; 97US-0048893.
 XX 06-JUN-1997; 97US-0048896.
 XX 06-JUN-1997; 97US-0048899.
 XX 06-JUN-1997; 97US-0048915.
 XX 06-JUN-1997; 97US-0048949.
 XX 06-JUN-1997; 97US-0048964.
 XX 06-JUN-1997; 97US-0048972.
 XX 06-JUN-1997; 97US-0049020.
 XX 06-JUN-1997; 97US-0049375.
 XX 05-SEP-1997; 97US-0057628.
 XX 05-SEP-1997; 97US-0057635.
 XX 05-SEP-1997; 97US-0057644.
 XX 05-SEP-1997; 97US-0057647.
 XX 05-SEP-1997; 97US-0057650.
 XX 05-SEP-1997; 97US-0057661.
 XX 05-SEP-1997; 97US-0057667.
 XX 05-SEP-1997; 97US-0057761.
 XX 05-SEP-1997; 97US-0057764.
 XX 05-SEP-1997; 97US-0057770.
 XX 05-SEP-1997; 97US-0057775.
 XX 05-SEP-1997; 97US-0057778.
 XX 06-JUN-1997; 97US-0048875.
 XX 06-JUN-1997; 97US-0048878.
 XX 06-JUN-1997; 97US-0048882.
 XX 06-JUN-1997; 97US-0048885.
 XX 06-JUN-1997; 97US-0048894.
 XX 06-JUN-1997; 97US-0048897.
 XX 06-JUN-1997; 97US-0048900.
 XX 06-JUN-1997; 97US-0048916.
 XX 06-JUN-1997; 97US-0048962.
 XX 06-JUN-1997; 97US-0048970.
 XX 06-JUN-1997; 97US-0048974.
 XX 06-JUN-1997; 97US-0049373.
 XX 05-SEP-1997; 97US-0057584.
 XX 05-SEP-1997; 97US-0057629.
 XX 05-SEP-1997; 97US-0057642.
 XX 05-SEP-1997; 97US-0057645.
 XX 05-SEP-1997; 97US-0057648.
 XX 05-SEP-1997; 97US-0057651.
 XX 05-SEP-1997; 97US-0057668.
 XX 05-SEP-1997; 97US-0057668.
 XX 05-SEP-1997; 97US-0057762.
 XX 05-SEP-1997; 97US-0057765.
 XX 05-SEP-1997; 97US-0057771.
 XX 05-SEP-1997; 97US-0057776.
 XX 06-JUN-1997; 97US-0048876.
 XX 06-JUN-1997; 97US-0048880.
 XX 06-JUN-1997; 97US-0048883.
 XX 06-JUN-1997; 97US-0048892.
 XX 06-JUN-1997; 97US-0048895.
 XX 06-JUN-1997; 97US-0048898.

us-09-380-826a-2.rng

Wed Mar 28 14:03:15 2001

```

PR 06-JUN-1997; 97US-0048901.
PR 06-JUN-1997; 97US-0048917.
PR 06-JUN-1997; 97US-0048963.
PR 06-JUN-1997; 97US-0048971.
PR 06-JUN-1997; 97US-0049019.
PR 06-JUN-1997; 97US-0049374.
PR 06-JUN-1997; 97US-0056227.
PR 05-SEP-1997; 97US-0057634.
PR 05-SEP-1997; 97US-0057643.
PR 05-SEP-1997; 97US-0057646.
PR 05-SEP-1997; 97US-0057649.
PR 05-SEP-1997; 97US-0057654.
PR 05-SEP-1997; 97US-0057666.
PR 05-SEP-1997; 97US-0057760.
PR 05-SEP-1997; 97US-0057763.
PR 05-SEP-1997; 97US-0057769.
PR 05-SEP-1997; 97US-0057774.
PR 05-SEP-1997; 97US-0057777.
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX Brewer LA, Carter KC, Ebner R, Endress GA;
XX Fan P, Feng P, Ferrie AM, Fischer CL, Florence C;
XX Florence K, Greene JM, Hu J, Kyaw H, Lafleur DW;
XX Li Y, Moore PA, Ni J, Olsen HS, Rosen CA, Ruben SM;
XX Shi Y, Soppet DR, Wei Y, Young P, Yu G, Zeng Z;
XX WPI: 1999-059865/05.
XX P-PSDB; W88721.
DR
XX New isolated human genes and the secreted polypeptides they encode -
PT useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders
PT
XX Claim 4: Page 450-451; 772pp: English.
PS
XX The invention relates to nucleic acid sequences (V84411 to V84633)
XX encoding human secreted proteins (W88534 to W88756). The secreted protein
XX gene sequences are deposited with the ATCC under deposit numbers ATCC
XX 97979, 97974, 97975, 97976, 97977, 209007, 209008, 209009, 209010,
XX 209011, 209080, 209081, 209082, 209083, 209084, 209085, 209511. Host
XX cells comprising recombinant vectors containing the nucleic acid
XX sequences are used for the recombinant production of the secreted
XX proteins. The polynucleotide and amino acid sequences are useful for
XX by protein or gene therapy, treating or ameliorating medical conditions e.g.
XX by determining the amount of the new polypeptides in a sample or by
XX determining the presence of mutations in the new polynucleotides.
XX Specific uses are described for each of the polynucleotides, based on
XX which tissues they are most highly expressed in, and include developing
XX products for the diagnosis or treatment of cancer, neurodegenerative
XX disorders, developmental abnormalities and foetal deficiencies, blood
XX disorders, tumours, leukemias, diseases of the immune system, autoimmune
XX diseases, hepatic and renal disease, lymphomas, inflammation, allergies,
XX ischemic shock, Alzheimer's and cognitive disorders involving osteoclasts
XX restenosis, prostate diseases, obesity, disorders involving osteoclasts
XX such as osteoporosis, arthritis or malignancies, infections of testes, lung
XX or thymus digestive/endocrine disorders, infections and AIDS. The
XX polypeptides are also useful for identifying their binding partners.
XX The present sequence represents a gene encoding a human secreted protein
XX (see descriptor line for gene number and clone identification).
XX
XX Sequence 951 BP; 296 A; 134 C; 192 G; 329 T; 0 other;
XX
XX Query Match 73.6%; Score 16.2; DB 20; Length 951;
XX Best Local Similarity 85.7%; Pred. No. 47;
XX Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
XX QY 1 tgttgatcacagaattgattg 21
XX ||||| || ||||| |||||
XX DB 524 TGTGGTCTCAAGATCTCAT 504

```

```

RESULT 6
V59636/c
ID V59636 standard; DNA; 1291 BP.
XX
XX AC V59636;
XX
XX DT 19-JAN-1999 (first entry)
XX
XX DE Human secreted protein gene 126 clone HELGH31.
XX
XX KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
XX diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
XX developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
XX immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
XX inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
XX cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
XX osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
XX endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX
XX OS Homo sapiens.
XX
XX PN WO9839448-A2.
XX
XX PD 11-SEP-1998.
XX
XX PF 06-MAR-1998; 98WO-US04493.
XX
XX 02-OCT-1997; 97US-0061060.
XX 07-MAR-1997; 97US-0038621.
XX 07-MAR-1997; 97US-0040161.
XX 07-MAR-1997; 97US-0040162.
XX 07-MAR-1997; 97US-0040163.
XX 07-MAR-1997; 97US-0040333.
XX 07-MAR-1997; 97US-0040334.
XX 07-MAR-1997; 97US-0040336.
XX 07-MAR-1997; 97US-0040626.
XX 11-APR-1997; 97US-0043311.
XX 11-APR-1997; 97US-0043312.
XX 11-APR-1997; 97US-0043313.
XX 11-APR-1997; 97US-0043314.
XX 11-APR-1997; 97US-0043568.
XX 11-APR-1997; 97US-0043569.
XX 11-APR-1997; 97US-0043576.
XX 11-APR-1997; 97US-0043578.
XX 11-APR-1997; 97US-0043580.
XX 11-APR-1997; 97US-0043669.
XX 11-APR-1997; 97US-0043670.
XX 11-APR-1997; 97US-0043671.
XX 11-APR-1997; 97US-0043672.
XX 23-MAY-1997; 97US-0043674.
XX 23-MAY-1997; 97US-0047492.
XX 23-MAY-1997; 97US-0047500.
XX 23-MAY-1997; 97US-0047501.
XX 23-MAY-1997; 97US-0047502.
XX 23-MAY-1997; 97US-0047503.
XX 23-MAY-1997; 97US-0047581.
XX 23-MAY-1997; 97US-0047582.
XX 23-MAY-1997; 97US-0047583.
XX 23-MAY-1997; 97US-0047584.
XX 23-MAY-1997; 97US-0047585.
XX 23-MAY-1997; 97US-0047586.
XX 23-MAY-1997; 97US-0047587.
XX 23-MAY-1997; 97US-0047588.
XX 23-MAY-1997; 97US-0047589.
XX 23-MAY-1997; 97US-0047590.
XX 23-MAY-1997; 97US-0047592.
XX 23-MAY-1997; 97US-0047593.
XX 23-MAY-1997; 97US-0047594.
XX 23-MAY-1997; 97US-0047595.
XX 23-MAY-1997; 97US-0047596.
XX 23-MAY-1997; 97US-0047597.
XX 23-MAY-1997; 97US-0047598.

```


PR 23-MAY-1997; 97US-0047599.
 PR 23-MAY-1997; 97US-0047600.
 PR 23-MAY-1997; 97US-0047601.
 PR 23-MAY-1997; 97US-0047612.
 PR 23-MAY-1997; 97US-0047613.
 PR 23-MAY-1997; 97US-0047614.
 PR 23-MAY-1997; 97US-0047615.
 PR 23-MAY-1997; 97US-0047617.
 PR 23-MAY-1997; 97US-0047618.
 PR 23-MAY-1997; 97US-0047632.
 PR 23-MAY-1997; 97US-0047633.
 PR 06-JUN-1997; 97US-0048964.
 PR 06-JUN-1997; 97US-0048974.
 PR 13-JUN-1997; 97US-0049610.
 PR 08-JUL-1997; 97US-0051926.
 PR 16-JUL-1997; 97US-0052874.
 PR 18-AUG-1997; 97US-0055724.
 PR 22-AUG-1997; 97US-0056630.
 PR 22-AUG-1997; 97US-0056631.
 PR 22-AUG-1997; 97US-0056632.
 PR 22-AUG-1997; 97US-0056636.
 PR 22-AUG-1997; 97US-0056637.
 PR 22-AUG-1997; 97US-0056662.
 PR 22-AUG-1997; 97US-0056664.
 PR 22-AUG-1997; 97US-0056845.
 PR 22-AUG-1997; 97US-0056862.
 PR 22-AUG-1997; 97US-0056864.
 PR 22-AUG-1997; 97US-0056872.
 PR 22-AUG-1997; 97US-0056874.
 PR 22-AUG-1997; 97US-0056875.
 PR 22-AUG-1997; 97US-0056876.
 PR 22-AUG-1997; 97US-0056877.
 PR 22-AUG-1997; 97US-0056878.
 PR 22-AUG-1997; 97US-0056879.
 PR 22-AUG-1997; 97US-0056880.
 PR 22-AUG-1997; 97US-0056881.
 PR 22-AUG-1997; 97US-0056882.
 PR 22-AUG-1997; 97US-0056884.
 PR 22-AUG-1997; 97US-0056886.
 PR 22-AUG-1997; 97US-0056887.
 PR 22-AUG-1997; 97US-0056888.
 PR 22-AUG-1997; 97US-0056889.
 PR 22-AUG-1997; 97US-0056892.
 PR 22-AUG-1997; 97US-0056893.
 PR 22-AUG-1997; 97US-0056894.
 PR 22-AUG-1997; 97US-0056903.
 PR 22-AUG-1997; 97US-0056908.
 PR 22-AUG-1997; 97US-0056909.
 PR 22-AUG-1997; 97US-0056910.
 PR 22-AUG-1997; 97US-0056911.
 PR 05-SEP-1997; 97US-0057650.
 PR 05-SEP-1997; 97US-0057659.
 PR 05-SEP-1997; 97US-0057761.
 PR 12-SEP-1997; 97US-0058785.
 PR (HUMA-) HUMAN GENOME SCI INC.
 PA Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress CA;
 PI Feng P, Fertie AM, Fischer CL, Florence KA, Greene JM, Hu JS;
 PI Kyaw H, Lafleur DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA;
 PI Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;
 XX WPI; 1998-506364/43.
 DR P-PSDB; W74854.
 XX
 PA New isolated human genes and the secreted polypeptide(s) they encode
 PT - useful for diagnosis and treatment of e.g. cancers, neurological
 PT disorders, immune diseases, inflammation of blood disorders
 XX
 PS Claim 1; Page 360-361; 721pp; English.
 XX
 CC This sequence represents a nucleic acid molecule designated Gene 126 from
 CC the human CDNA clone HELGH31 (deposited as clone ATCC 97902 and ATCC

CC 209048) which encodes a secreted human protein. The gene can be used to
 CC generate fusion proteins by linking to the gene to a human immunoglobulin
 CC Fc portion (e.g. V59502) for increasing the stability of the fused
 CC protein as compared to the human protein only.
 CC The invention relates to 186 novel genes and their fragments (nucleic
 CC acid sequences: V59511-V59812; amino acid sequences W74731-W75026) which
 CC are useful for preventing, treating or ameliorating medical conditions
 CC e.g. by protein or gene therapy. Also, pathological conditions can be
 CC diagnosed by determining the amount of the new polypeptides in a sample
 CC or by determining the presence of mutations in the new polynucleotides.
 CC Specific uses are described for each of the 186 polynucleotides, based on
 CC which tissues they are most highly expressed in (see V59511 for described
 CC uses).
 XX
 SQ Sequence 1291 BP; 391 A; 228 C; 203 G; 469 T; 0 other;

Query Match 73.6%; Score 16.2; DB 19; Length 1291;
 Best Local Similarity 85.7%; Pred. No. 48;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 tgttggtacacagatttgat 21
 ||| ||||| |||||
 Db 367 TGTATGATCACATGATTGAT 347

RESULT 7
 X61757/C
 ID X61757 standard; DNA; 1473 BP.
 XX
 AC X61757;
 XX
 DT 19-JUL-1999 (first entry)
 XX
 DE B. burgdorferi antigenic protein coding sequence, f893.nt.
 KW Antigenic protein; vaccine; Lyme disease; infection; detection; ss.
 XX
 OS Borrelia burgdorferi.
 XX
 PN WO9859071-A1.
 XX
 PD 30-DEC-1998.
 XX
 PF 18-JUN-1998; 98WO-US12718.
 XX
 PR 03-SEP-1997; 97US-0057483.
 PR 20-JUN-1997; 97US-0050359.
 PR 22-JUL-1997; 97US-0053344.
 PR 22-JUL-1997; 97US-0053377.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (MEDI-) MEDIMMUNE INC.
 XX
 PI Choi GH, Erwin AL, Hanson MS, Lathigra R;
 XX
 DR WPI; 1999-189980/16.
 DR P-PSDB; Y20060.
 XX
 PT New isolated Borrelia burgdorferi nucleic acids - used to develop
 PT products for the diagnosis, prevention and treatment of diseases
 PT caused by Borrelia, particularly Lyme disease
 XX
 PS Claim 1; Page 180; 275pp; English.
 XX
 CC This sequence encodes a Borrelia burgdorferi (Bb) protein of the
 CC invention, which is suitable for use in a vaccine. The Bb polypeptides
 CC can be used in vaccines for eliciting protective antibodies to members of
 CC the Borrelia genus, particularly for the use against Lyme disease in
 CC humans and animals. They can be used for preventing or attenuating an
 CC infection caused by a member of the Borrelia genus. The products can also
 CC be used for detection of members of the Borrelia genus.

Oy 2 gttggatcacagaatttgata 22
 ||||| ||||| |||||
 Db 42 gttggatcacagaatttgta 62

RESULT 10

Z49816
 ID Z49816 standard; DNA; 2001 BP.

AC Z49816;

XX 18-APR-2000 (first entry)

DE EcoRI fragment of lamda clone #15-14 with maize ACCase gene portion.

XX Herbicide resistance; gene modification; lamda clone #15-14;

KW maize acetyl CoA carboxylase; ACCase; herbicide; plant oil content;

KW marker-assisted plant selection; groat oil trait;

KW restriction fragment length polymorphism mapping;

XX high-energy animal feed; low-fat human food; ds.

OS Zea mays.

XX WO9967367-A1.

PN 29-DEC-1999.

PD 22-JUN-1999; 99WO-US14022.

PF 22-JUN-1998; 98US-0090240.

PR 02-JUL-1998; 98US-0091640.

XX (MINU) UNIV MINNESOTA.

PA (USDA) US DEPT OF AGRICULTURE.

PA (EGLI/) EGLI M A.

PA (GROH/) GROH S.

PA (KIAN/) KIANIAN S F.

PA (PHIL/) PHILLIPS R L.

PA (RINE/) RINES H W.

PA (SOME/) SOMERS D A.

XX Egli MA, Groh S, Kianian SF, Phillips RL, Rines HW, Somers DA;

PI WPI; 2000-147205/13.

XX New DNA encoding acetyl-CoA carboxylase from oats, used to produce transformed plants with herbicide resistance and altered oil content

PT Disclosure; Fig 10; 197pp; English.

PS The present DNA sequence is a 2kb EcoRI fragment of lamda clone #15-14 including a portion of a maize acetyl CoA carboxylase (ACCase) gene located at bases 2883 to 83 from the 3' stop codon. ACCase is involved in fatty acid synthesis and is the target of the specified herbicides. Transformation of plants with ACCase imparts resistance to cyclohexanedione and aryloxyphenoxypyranoic acid herbicides and alter the oil content. The ACCase DNA is also used as source of probes and primers for the identification of transgenic plants; in marker-assisted plant selection and for restriction fragment length polymorphism mapping, used for high-energy animal feed and high-fiber, low-fat human food and in genetic dissection of the groat oil trait.

XX Sequence 2001 BP; 546 A; 406 C; 494 G; 554 T; 1 other;

Query Match

Best Local Similarity 73.6%; Score 16.2; DB 21; Length 2001;

Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 2 gttggatcacagaatttgata 22

||||| ||||| |||||

Db 42 gttggatcacagaatttgta 62

RESULT 11

T64683
 ID T64683 standard; DNA; 3822 BP.

XX T64683;

XX 17-MAR-1998 (first entry)

XX M. leprae gyrA precursor coding sequence.

DE Mycobacterium sp.; internal sequence; intein; immature; gyrase; protein splicing; precursor; gyrA; ss.

OS Mycobacterium leprae.

XX FR2739859-A1.

PN 18-APR-1997.

PD 17-OCT-1995; 95FR-0012162.

PF 17-OCT-1995; 95FR-0012162.

PR 17-OCT-1995; 95FR-0012162.

XX (INSP) INST PASTEUR.

PA Cole S, Fsihi H;

PI WPI; 1997-247976/23.

DR P-PSDB; W15078.

XX Mycobacterial DNA gyrase precursor protein - and polypeptide(s) corresponding to mycobacterial DNA gyrase intein sequences

PT Claim 5; Pages 33-34; 55pp; French.

PS This DNA, isolated from Mycobacterium leprae, encodes a precursor (immature) gyrase protein from which an "intein" (see W15074) is excised during maturation. The gyrase and its coding sequence are useful to fabricate products that alter the maturation of proteins essential for the development of infectious agents by altering their protein splicing of precursor polypeptides of the proteins. The inteins encoded by the gyrA genes of Mycobacterium leprae, M. flavescens, M. goodii and M. kansasii are used: (a) to modify the genome of a eukaryotic cell that lacks endogenous biological activity identical to that of the polypeptide; (b) to replace a copy of a gene present in a recipient genome by integration of a gene different from that where the integration takes place; and (c) for targeted insertion of a foreign DNA sequence into a selected site in the genome of a eukaryotic cell not containing the specific cleavage sites of the polypeptide.

XX Query Match 73.6%; Score 16.2; DB 18; Length 3822;

Best Local Similarity 85.7%; Pred. No. 54;

Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 2 gttggatcacagaatttgata 22

||||| ||||| |||||

Db 2334 gttggatcacagaatttgta 2354

RESULT 12

Q42933
 ID Q42933 standard; DNA; 4346 BP.

XX Q42933;

DT 15-OCT-1993 (first entry)

XX A3 maize ACCase cDNA clone.

Wed Mar 28 14:03:15 2001

XX Acetyl CoA carboxylase; ACCase; probe: expression; oil content;
 KW lower; modified; increased; oilseed rape; sunflower; resistant;
 KW resistance; grass-weed herbicides; fluzafop; alkyketones; maize;
 KW wheat; barley; ss.
 XX
 OS Zea mays.
 XX
 FH Key Location/Qualifiers
 FT CDS 3..3947
 FT /*tag= a
 FT /note= "ACCase"
 FT 1875..1878
 FT misc_feature
 FT /*tag= b
 FT /note= "GAN encodes Glu or Asp"
 XX
 PN W09311243-A.
 XX
 PD 10-JUN-1993.
 XX
 PF 27-NOV-1992; 92WO-GB02205.
 XX
 PR 28-NOV-1991; 91GB-0025330.
 XX
 PA (ICIL) ICI AUSTRALIA OPERATIONS PTY LTD.
 PA (ICIL) IMPERIAL CHEM IND PLC.
 XX
 PI Ashton AR, Jenkins CL, Whitfield PR;
 XX
 DR WPI; 1993-197061/24.
 DR P-PSDB; R36781.
 XX
 PT DNA clones for use in probing plant DNA - comprise maize acetyl
 PT coenzyme-A carboxylase gene flanked by heterologous DNA
 XX
 PS Claim 1; Fig 3; 69pp; English.
 XX
 CC The sequence is that of the A3 acetyl CoA carboxylase (ACCase) cDNA
 CC maize clone which can be used to probe plant DNA to isolate other
 CC such clones. It may be used with plant regulatory sequences to
 CC produce expression cassettes for ACCase. These cassettes are used to
 CC transform plants to downregulate prodn. of ACCase. This can alter
 CC the composition of seeds or other plant parts, e.g. enabling oil-
 CC bearing plants such as oilseed rape, sunflower or oilpalm, having a
 CC lower or modified oil content to be produced. It can be used to form
 CC expression cassettes for overexpression of ACCase., leading to prodn.
 CC of plants with an increased oil content. It may also be used to recover
 CC the ACCase gene promoter. This can be used to generate RNA in a
 CC tissue-specific or developmentally regulated manner, and this RNA
 CC used to inhibit ACCase expression. Monocotyledonous plants can be made
 CC resistant to grass-weed herbicides such as fluzafop and alkyketones
 CC by transforming with cassettes adapted to express ACCase. This is
 CC achieved by overexpression of monocot ACCase. expression of dicot
 CC ACCase (which is relatively tolerant to these herbicides), or
 CC possibly by expression of a resistant form of maize ACCase. This
 CC provides crop plants which are resistant to herbicides used to
 CC destroy weeds growing between the plants. This allows the overall
 CC applicn. of herbicide without affecting the crop, and is also useful
 CC where there has been short term carryover of herbicide from the
 CC previous crop. Using the method, weeds such as wild oats may be
 CC controlled, and the farmer is given extra options to obtain an improved
 CC harvest using means which are safer, cheaper or more effective.
 CC Suitable crops are maize, wheat and barley.
 XX
 SQ Sequence 4346 BP; 1239 A; 829 C; 1064 G; 1213 T; 1 other;

Query Match 73.6%; Score 16.2; DB 14; Length 4346;
 Best Local Similarity 85.7%; Pred. No. 55;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 2 gttggatcacagaattgtgata 22
 ||||||| ||||| ||| ||

Db 902 gttggatgacagaagttgtta 922
 RESULT 13
 T39905
 ID T39905 standard; cDNA; 7470 BP.
 XX
 AC T39905;
 XX
 DT 23-JAN-1997 (first entry)
 XX
 DE Maize acetyl CoA carboxylase cDNA.
 XX
 KW Acetyl CoA carboxylase; ACCase; herbicide tolerance;
 KW cyclohexanedione; aryloxyphenoxypionic acid; vegetable oil;
 KW oilseed; maize; corn; ss.
 XX
 OS Zea mays line A188.
 XX
 FH Key Location/Qualifiers
 FT CDS 37..7014
 FT /*tag= a
 FT /EC_number= 6.4.1.2
 FT
 XX W09631609-A2.
 PN
 XX 10-OCT-1996.
 XX
 PD 04-APR-1996; 96WO-US04625.
 XX
 PF 05-APR-1995; 95US-0417089.
 XX
 PR (MINU) UNIV MINNESOTA.
 XX
 PA Egli MA, Gengenbach BG, Gronwald JW, Lutz SM, Somers DA;
 PI Wyse DL;
 PI
 XX WPI; 1996-465030/46.
 DR P-PSDB; W05590.
 DR
 XX DNA encoding maize acetyl coenzyme A carboxylase gene - used for
 XX prodn. of plants with herbicide tolerance or altered oil content
 PT
 PT Claim 2; Page 78-80; 131pp; English.
 XX
 PS A cDNA sequence (T39905) codes for maize acetyl CoA carboxylase
 CC (ACCase) (W05590), an enzyme that plays a central role in fatty
 CC acid biosynthesis and accumulation in plants and seeds. The
 CC complete sequence was deduced from cDNA clone #18-5, obtd. from
 CC a lambda gt11 cDNA library of maize inbred A188 seedlings, and from
 CC PCR amplifications (see also T39918-23). Genomic ACCase DNA
 CC fragments have also been isolated (see also T39906-16). ACCase DNA
 CC can be incorporated into a vector and used to increase the
 CC herbicide tolerance or oil content of a transgenic plant, or used
 CC to produce ACCase in heterologous systems.
 CC
 XX Sequence 7470 BP; 2119 A; 1429 C; 1844 G; 2071 T; 7 other;
 SQ

Query Match 73.6%; Score 16.2; DB 17; Length 7470;
 Best Local Similarity 85.7%; Pred. No. 58;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 2 gttggatcacagaattgtgata 22
 ||||||| ||||| ||| ||

Db 3972 gttggatgacagaagttgtta 3992
 RESULT 14
 V29317
 ID V29317 standard; cDNA; 7470 BP.
 XX
 AC V29317;

XX 30-JUL-1998 (first entry)
 XX Maize ACCase enzyme encoding cDNA.
 DE ACCase; maize; herbicide resistant; corn plant; tolerance; Acc1; Acc2;
 KW cyclohexanedione; CHD; aryloxyphenoxypyranoic acid; APA; enzyme; ss.
 XX Zea mays.
 XX Key Location/Qualifiers
 XX CDS 37..7014
 XX /*tag= a
 XX /transl_except= (pos:4339..4341, aa: His)
 XX /transl_except= (pos:4456..4458, aa: Ser)
 XX /transl_except= (pos:4359..4461, aa: Lys)
 XX /product= "ACCase"
 XX W09808963-A1.
 XX 05-MAR-1998.
 XX 29-AUG-1997; 97WO-US15344.
 XX 30-AUG-1996; 96US-0697826.
 XX (MINU) UNIV MINNESOTA.
 XX Egli MA, Gengenbach BG, Lutz SM, Marshall LC, Parker WB;
 PI Somers DA, Vandee KL, Wyse DL;
 XX WPI: 1998-207043/18.
 XX P-PSDB: W56736.
 XX Herbicide resistant corn plants - prepared using Acc1 and Acc2 gene
 XX combinations
 XX Example 7; Fig 3A-E; 112pp; English.
 XX This cDNA encodes a maize ACCase enzyme. This can be used in a method
 CC of preparing an herbicide resistant corn plant which comprises crossing a
 CC first corn plant to a second corn plant so as to yield a progeny plant,
 CC where the first corn plant comprises at least 1 herbicide resistant
 CC allele and the second plant comprises at least 1 herbicide resistant
 CC allele which is not allelic to the herbicide resistant allele present in
 CC the first plant. The herbicide resistant corn plants are prepared using
 CC Acc1 and Acc2 gene combinations which impart cyclohexanedione (CHD) or
 CC aryloxyphenoxypyranoic acid (APA) herbicide tolerance to the corn plant.
 CC The methods are used to impart CHD and APA herbicide tolerance to corn
 CC plants and to produce CHD or APA herbicide tolerant Zea mays (corn)
 CC homozygous or heterozygous for Acc1 and homozygous or heterozygous for
 CC Acc2. The methods can also be used to impart tolerance to a corn plant
 CC to an agent which inhibits acetyl CoA carboxylase, selected from
 CC 3-(2,4-dichlorophenyl)-perhydroindolizine-2,4-dione, 3-isopropyl-6-
 CC (N-[2,2-dimethylpropyl]-acetamido)-1,3,5-triazine-2,4 (1H,3H)dione,
 CC sorafen A and their structural analogues.
 XX Sequence 7470 BP; 2119 A; 1429 C; 1844 G; 2071 T; 7 other;

Query Match 73.6%; Score 16.2; DB 19; Length 7470;
 Best Local Similarity 85.7%; Pred. No. 58;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 2 gttggatcacagaattgata 22
 Db 3972 gttggatgacagaagtgtgta 3992

RESULT 15
 ID 249820
 XX 249820 standard; cDNA; 7470 BP.

AC 249820;
 XX 18-APR-2000 (first entry)
 XX Maize acetyl CoA carboxylase cDNA.
 DE Herbicide resistance; gene modification;
 KW maize acetyl CoA carboxylase; ACCase; herbicide; plant oil content;
 KW marker-assisted plant selection; groat oil trait;
 KW restriction fragment length polymorphism mapping;
 KW high-energy animal feed; low-fat human food; ss.
 XX Zea mays.
 XX Key Location/Qualifiers
 XX CDS 37..7014
 XX /*tag= a
 XX /product= "Maize acetyl CoA carboxylase"
 XX 1..36
 XX /*tag= b
 XX 7014..7470
 XX /*tag= c
 XX W09967367-A1.
 XX 29-DEC-1999.
 XX 22-JUN-1999; 99WO-US14022.
 XX 22-JUN-1998; 98US-0090240.
 XX 02-JUL-1998; 98US-0091640.
 XX (MINU) UNIV MINNESOTA.
 XX (USDA) US DEPT OF AGRICULTURE.
 XX (EGLI/) EGLI M A.
 XX (GROH/) GROH S.
 XX (KIAN/) KIANIAN S F.
 XX (PHIL/) PHILLIPS R L.
 XX (RINE/) RINES H W.
 XX (SOME/) SOMERS D A.
 XX Egli MA, Groh S, Kianian SF, Phillips RL, Rines HW, Somers DA;
 WPI: 2000-147205/13.
 XX P-PSDB: Y44687.
 XX New DNA encoding acetyl-CoA carboxylase from oats, used to produce
 XX transformed plants with herbicide resistance and altered oil content
 XX Example 6; Fig 13; 197pp; English.
 XX The present cDNA sequence encodes maize acetyl CoA carboxylase (ACCase).
 CC ACCase is involved in fatty acid synthesis and is the target of
 CC the specified herbicides. Transformation of plants with ACCase imparts
 CC resistance to cyclohexanedione and aryloxyphenoxypyranoic acid
 CC herbicides and alter the oil content. The ACCase cDNA is also used as
 CC source of probes and primers for the identification of transgenic
 CC plants; in marker-assisted plant selection and for restriction fragment
 CC length polymorphism mapping, used for high-energy animal feed and
 CC high-fiber, low-fat human food and in genetic dissection of the groat
 CC oil trait.
 XX Sequence 7470 BP; 2119 A; 1430 C; 1843 G; 2071 T; 7 other;

Query Match 73.6%; Score 16.2; DB 21; Length 7470;
 Best Local Similarity 85.7%; Pred. No. 58;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 2 gttggatcacagaattgata 22
 Db 3972 gttggatgacagaagtgtgta 3992

us-09-380-826a-2.rng

Wed Mar 28 14:03:15 2001

Search completed: March 27, 2001, 08:22:58
Job Time: 5300 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 27, 2001, 08:19:18 ; Search time 132.3 Seconds
(without alignments)
26.799 Million cell updates/sec

Title: US-09-380-826A-2

Perfect score: 22

Sequence: 1 ttttgatcacaaagtattgata 22

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 280836 seqs, 80580151 residues

Total number of hits satisfying chosen parameters: 561672

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:*

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
- 3: /cgn2_6/ptodata/2/ina/6_COMB.seq:*
- 4: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
- 5: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	16.2	73.6	2000	1	US-08-014-326-1
2	16.2	73.6	2001	3	US-08-417-089-1
3	16.2	73.6	2001	3	US-08-695-651-1
4	16.2	73.6	4345	2	US-08-244-537-1
5	16.2	73.6	7470	3	US-08-417-089-5
6	16.2	73.6	7470	3	US-08-695-651-5
7	15.6	70.9	296	3	US-08-602-145-14
8	15.6	70.9	1096	1	US-08-684-862-8
9	15.6	70.9	1989	2	US-08-792-055-1
10	15.6	70.9	7745	1	US-08-299-675-1
11	15.6	70.9	7745	1	US-08-485-241-1
12	15.6	70.9	7745	2	US-08-874-162-1
13	15.2	69.1	1016	1	US-08-399-986B-3
14	15.2	69.1	1016	1	US-08-493-754A-3
15	15.2	69.1	2182	1	US-08-399-986B-1
16	15.2	69.1	2182	1	US-08-493-754A-1
17	15.2	69.1	2598	3	US-08-745-892-20
18	15.2	69.1	5420	5	5256642-3
19	15.2	69.1	5420	5	5472939-3
20	15.2	69.1	6951	5	5256642-1
21	15.2	69.1	6951	5	5472939-1
22	14.8	67.3	1855	3	US-08-961-083-71
23	14.8	67.3	3159	1	US-08-119-361-4
24	14.8	67.3	3159	3	US-08-336-308A-3
25	14.8	67.3	3159	3	US-08-822-324-3
26	14.8	67.3	7266	3	US-08-336-308A-9
27	14.8	67.3	7266	3	US-08-822-324-5
28	14.8	67.3	8640	1	US-08-570-311-28

29 14.8 67.3 9919 3 US-08-880-179-1 Sequence 1, Appli
30 14.6 66.4 293 3 US-08-866-340-13 Sequence 13, Appli
31 14.6 66.4 1404 1 US-08-204-656B-1 Sequence 1, Appli
32 14.6 66.4 1404 1 US-08-204-656B-3 Sequence 3, Appli
33 14.6 66.4 1404 1 US-08-204-656B-5 Sequence 5, Appli
34 14.6 66.4 1404 1 US-08-204-656B-7 Sequence 7, Appli
35 14.6 66.4 1404 1 US-08-470-702-1 Sequence 1, Appli
36 14.6 66.4 1404 1 US-08-470-702-2 Sequence 2, Appli
37 14.6 66.4 1404 1 US-08-470-702-3 Sequence 3, Appli
38 14.6 66.4 1404 1 US-08-470-702-4 Sequence 4, Appli
39 14.6 66.4 1404 1 US-08-467-831-1 Sequence 1, Appli
40 14.6 66.4 1404 1 US-08-467-831-2 Sequence 2, Appli
41 14.6 66.4 1404 1 US-08-467-831-3 Sequence 3, Appli
42 14.6 66.4 1404 1 US-08-467-831-4 Sequence 4, Appli
C 43 14.6 66.4 1608 2 US-08-424-224-1 Sequence 1, Appli
C 44 14.6 66.4 1608 4 PCT-US94-02891-68 Sequence 68, Appli
45 14.6 66.4 12047 2 US-09-022-461-1 Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-014-326-1
; Sequence 1, Application US/08014326
; Patent No. 5498544
; GENERAL INFORMATION:
; APPLICANT: Gengenbach, Burtle G.
; APPLICANT: Somers, David A.
; APPLICANT: Wyse, Donald L.
; APPLICANT: Gronwald, John W.
; APPLICANT: Egli, Margaret A.
; APPLICANT: Lutz, Shiela M.
; TITLE OF INVENTION: Method and An Acetyl CoA Carboxylase
; TITLE OF INVENTION: Alteration in Oil Content of Plants
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant & Gould
; STREET: 3100 No. 5498544west Center
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/014,326
; FILING DATE: 05-FEB-1993
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/917,462
; FILING DATE: 21-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/538,674
; FILING DATE: 18-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/269,584
; FILING DATE: 10-NOV-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Kowalchuk, Katherine M.
; REGISTRATION NUMBER: 36,848
; REFERENCE/DOCKET NUMBER: 600.258-US01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-332-5300
; TELEFAX: 612-332-9081
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2000 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

```
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; IMMEDIATE SOURCE:
; CLONE: 2 kb fragment of lambda clone #15-14
; US-08-014-326-1

Query Match      73.6%; Score 16.2; DB 1; Length 2000;
Best Local Similarity 85.7%; Pred. No. 20;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 gttggtacacaagatttgata 22
   ||||| ||||| ||||| |||||
Db 42 GTTGATGACAAAGAGTTGTGA 62

RESULT 2
US-08-417-089-1
; Sequence 1, Application US/08417089
; Patent No. 6089298
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: METHODS AND AN ACETYL COA CARBOXYLASE GENE
; TITLE OF INVENTION: FOR CONFERRING HERBICIDE TOLERANCE AND AN ALTERATION IN
; TITLE OF INVENTION: OIL CONTENT OF PLANTS
; NUMBER OF SEQUENCES: 11
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/417,089
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2001 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-417-089-1

Query Match      73.6%; Score 16.2; DB 3; Length 2001;
Best Local Similarity 85.7%; Pred. No. 20;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 gttggtacacaagatttgata 22
   ||||| ||||| ||||| |||||
Db 42 GTTGATGACAAAGAGTTGTGA 62

RESULT 3
US-08-695-651-1
; Sequence 1, Application US/08695651
; Patent No. 6146867
; GENERAL INFORMATION:
; APPLICANT: Gengenbach, B. G.
; APPLICANT: Somers, D. A.
; APPLICANT: Wyse, D. L.
; APPLICANT: Gronwald, J. W.
; APPLICANT: Egli, M. A.
; APPLICANT: Lutz, S. M.
; TITLE OF INVENTION: METHOD FOR PRODUCING PLANT ACETYL COA
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schwegman, Lundberg, Woessner & Kluth, P.A.
; STREET: P.O. Box 2938
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
```

```
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/695,651
; FILING DATE: 12-AUG-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/417089
; FILING DATE: 05-APR-1995
; APPLICATION NUMBER: 08/014326
; FILING DATE: 05-FEB-1993
; APPLICATION NUMBER: 07/917462
; FILING DATE: 21-JUL-1992
; APPLICATION NUMBER: 07/538674
; FILING DATE: 18-JUN-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Woessner, Warren D
; REGISTRATION NUMBER: 30,440
; REFERENCE/DOCKET NUMBER: 600.318US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-373-6900
; TELEFAX: 612-339-3061
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2001 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; US-08-695-651-1

Query Match      73.6%; Score 16.2; DB 3; Length 2001;
Best Local Similarity 85.7%; Pred. No. 20;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 gttggtacacaagatttgata 22
   ||||| ||||| ||||| |||||
Db 42 GTTGATGACAAAGAGTTGTGA 62

RESULT 4
US-08-244-537-1
; Sequence 1, Application US/08244537
; Patent No. 5854420
; GENERAL INFORMATION:
; APPLICANT: ASHTON, ANTHONY R.
; APPLICANT: JENKINS, COLIN L.D.
; APPLICANT: WHITFIELD, PAUL R.
; TITLE OF INVENTION: MAIZE ACETYL COA CARBOXYLASE ENCODING
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: D. C.
; COUNTRY: U.S.A.
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/244,537
; FILING DATE: 18-AUG-1994
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9125330.2
```


;; FILING DATE: 28-NOV-1991
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: PCT/GB92/02205
;; FILING DATE: 27-NOV-1992
;; ATTORNEY/AGENT INFORMATION:
;; NAME: KOKULIS, PAUL N.
;; REGISTRATION NUMBER: 16,773
;; REFERENCE/DOCKET NUMBER: 203094/SEE 36663/UST
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (202) 861-3000
;; TELEFAX: (202) 822-0944
;; TELEX: 6714627 CUSH
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 4345 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: both
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
US-08-244-537-1

Query Match 73.6%; Score 16.2; DB 2; Length 4345;
Best Local Similarity 85.7%; Pred. No. 22;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 gttggatcacagaatttgata 22
||||| ||||| ||| ||
Db 902 GTTGGATGACAAGAGTGTGTTA 922

RESULT 5
US-08-417-089-5
; Sequence 5, Application US/08417089
; Patent No. 6069298
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: METHODS AND AN ACETYL COA CARBOXYLASE GENE
; TITLE OF INVENTION: FOR CONFERRING HERBICIDE TOLERANCE AND AN ALTERATION IN
; TITLE OF INVENTION: OIL CONTENT OF PLANTS
; NUMBER OF SEQUENCES: 11
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/417,089
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7470 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-417-089-5

Query Match 73.6%; Score 16.2; DB 3; Length 7470;
Best Local Similarity 85.7%; Pred. No. 23;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 gttggatcacagaatttgata 22
||||| ||||| ||| ||
Db 3972 GTTGGATGACAAGAGTGTGTTA 3992

RESULT 6
US-08-695-651-5
; Sequence 5, Application US/08695651
; Patent No. 6146867
; GENERAL INFORMATION:
; APPLICANT: Gengenbach, B. G.

;; APPLICANT: Somers, D. A.
;; APPLICANT: Wyse, D. L.
;; APPLICANT: Gronwald, J. W.
;; APPLICANT: Egli, M. A.
;; APPLICANT: Lutz, S. M.
;; TITLE OF INVENTION: METHOD FOR PRODUCING PLANT ACETYL COA
;; NUMBER OF SEQUENCES: 11
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Schwegman, Lundberg, Woessner & Kluth, P.A.
;; STREET: P.O. Box 2938
;; CITY: Minneapolis
;; STATE: MN
;; COUNTRY: USA
;; ZIP: 55402
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette
;; COMPUTER: IBM Compatible
;; OPERATING SYSTEM: DOS
;; SOFTWARE: FastSEQ Version 2.0
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/695,651
;; FILING DATE: 12-AUG-1996
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/417089
;; FILING DATE: 05-APR-1995
;; APPLICATION NUMBER: 08/014326
;; FILING DATE: 05-FEB-1993
;; APPLICATION NUMBER: 07/917462
;; FILING DATE: 21-JUL-1992
;; APPLICATION NUMBER: 07/538674
;; FILING DATE: 18-JUN-1990
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Woessner, Warren D
;; REGISTRATION NUMBER: 30,440
;; REFERENCE/DOCKET NUMBER: 600.318US3
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 612-373-6900
;; TELEFAX: 612-339-3061
;; TELEX:
;; INFORMATION FOR SEQ ID NO: 5:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 7470 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: CDNA
US-08-695-651-5

Query Match 73.6%; Score 16.2; DB 3; Length 7470;
Best Local Similarity 85.7%; Pred. No. 23;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 gttggatcacagaatttgata 22
||||| ||||| ||| ||
Db 3972 GTTGGATGACAAGAGTGTGTTA 3992

RESULT 7
US-08-602-145-14/c
; Sequence 14, Application US/08602145
; Patent No. 6025336
; GENERAL INFORMATION:
; APPLICANT: Goltry, Kristin L.
; APPLICANT: Greenberger, Joel S.
; TITLE OF INVENTION: DETERMINING EXPOSURE TO IONIZING RADIATION
; TITLE OF INVENTION: AGENT WITH PERSISTENT BIOLOGICAL MARKERS
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington

Query Match 70.9%; Score 15.6; DB 1; Length 1096;
Best Local Similarity 81.8%; Pred. NO. 37;
Matches 18; Conservative 0; Mismatches 4; Indels 0

Qy 1 tgttgatcacaagatttgata 22
||| | ||||| ||||| ||
Db 112 TGAGTATCACAAAGGTTTGCTA 91

```

; TELEPHONE: (404) 688-0770
;
; TELEFAX: (404) 688-9880
;
; INFORMATION FOR SEQ ID NO: 1:
;     SEQUENCE CHARACTERISTICS:
;         LENGTH: 1989 base pairs
;         TYPE: nucleic acid
;         STRANDEDNESS: double
;         TOPOLOGY: linear
;     MOLECULE TYPE: DNA (genomic)
;     US-08:792-055-1

```

Query Match 70.9%; Score 15.6; DB 2; Length 1989;
Best Local Similarity 81.8%; Pred. No. 39;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 tgttgatcacaaagatttgata 22
||||| ||||| ||||| ||
Db 288 TGTGGATCAACAGATTGTA 267

RESULT 10
US-08-299-675-1/c
; Sequence 1, Application US/08299675
; Patent No. 5589618
; GENERAL INFORMATION:
; APPLICANT: Hannah, L. Curtis
; APPLICANT: Giroux, Michael
; TITLE OF INVENTION: Materials and Methods for Increasing
; TITLE OF INVENTION: Corn Seed Weight
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/299,675
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: UF146
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7745 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-299-675-1

Query Match 70.9%; Score 15.6; DB 1; Length 7745;
Best Local Similarity 81.8%; Pred. No. 46;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 tgttgatcacaaagatttgata 22
||||| ||||| ||||| ||
Db 3600 TGTGGATAACAAGATGACATA 3579

RESULT 11
US-08-485-241-1/c
; Sequence 1, Application US/08485241
; Patent No. 5650557
; GENERAL INFORMATION:
; APPLICANT: Hannah, L. Curtis
; APPLICANT: Giroux, Michael
; TITLE OF INVENTION: Materials and Methods for Increasing
; TITLE OF INVENTION: Corn Seed Weight
; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/485,241
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/299,675
; FILING DATE: 1-SEP-1994
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Pace, Doran R.
; REGISTRATION NUMBER: 38,261
; REFERENCE/DOCKET NUMBER: UF146.C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7745 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-485-241-1

Query Match 70.9%; Score 15.6; DB 1; Length 7745;
Best Local Similarity 81.8%; Pred. No. 46;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 tgttgatcacaaagatttgata 22
||||| ||||| ||||| ||
Db 3600 TGTGGATAACAAGATGACATA 3579

RESULT 12
US-08-874-162-1/c
; Sequence 1, Application US/08874162
; Patent No. 5872216
; GENERAL INFORMATION:
; APPLICANT: Hannah, L. Curtis
; APPLICANT: Giroux, Michael
; TITLE OF INVENTION: Materials and Methods for Increasing
; TITLE OF INVENTION: Corn Seed Weight
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/874,162
; FILING DATE: 13-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/485,241
FILING DATE: 7-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION NUMBER: US 08/299,675
FILING DATE: 1-SEP-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Pace, Doran R.
REGISTRATION NUMBER: 38,261
REFERENCE/DOCKET NUMBER: UF-146C1D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 352-375-8100
TELEFAX: 352-372-5800
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 7745 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-874-162-1

Query Match 70.9%; Score 15.6; DB 2; Length 7745;
Best Local Similarity 81.8%; Pred. No. 46;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 tggatcacacagatttgata 22
||||| ||||| ||||| |||||
Db 3600 TGTGGATAACAAGATGACATA 3579

RESULT 13
US-08-399-986B-3/C
Sequence 3, Application US/08399986B
Patent No. 5801041
GENERAL INFORMATION:
APPLICANT: Godwin, Andrew K.
TITLE OF INVENTION: No. 5801041el Gene Associated with Suppression
TITLE OF INVENTION: of Tumor Development
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman
STREET: 1601 Market Street
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/399,986B
FILING DATE: 06-MAR-1995
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Hagan, Patrick J.
REGISTRATION NUMBER: 27,643
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1016 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: not relevant
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO

US-08-399-986B-3

Query Match 69.1%; Score 15.2; DB 1; Length 1016;
Best Local Similarity 85.0%; Pred. No. 57;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 ttggatcacacagatttgata 22
||||| ||||| ||||| |||||
Db 830 TTGGTCTCAAGAAATTGATA 811

RESULT 14
US-08-493-754A-3/C
Sequence 3, Application US/08493754A
Patent No. 5821338
GENERAL INFORMATION:
APPLICANT: Godwin, Andrew K.
TITLE OF INVENTION: No. 5821338el Gene Associated with Suppression
TITLE OF INVENTION: of Tumor Development
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman
STREET: 1601 Market Street
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/493,754A
FILING DATE: 22-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Hagan, Patrick J.
REGISTRATION NUMBER: 27,643
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1016 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: not relevant
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-493-754A-3

Query Match 69.1%; Score 15.2; DB 1; Length 1016;
Best Local Similarity 85.0%; Pred. No. 57;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 ttggatcacacagatttgata 22
||||| ||||| ||||| |||||
Db 830 TTGGTCTCAAGAAATTGATA 811

RESULT 15
US-08-399-986B-1/C
Sequence 1, Application US/08399986B
Patent No. 5801041
GENERAL INFORMATION:
APPLICANT: Godwin, Andrew K.
TITLE OF INVENTION: No. 5801041el Gene Associated with Suppression
TITLE OF INVENTION: of Tumor Development
NUMBER OF SEQUENCES: 35

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
;; STREET: 1601 Market Street
;; CITY: Philadelphia
;; STATE: PA
;; COUNTRY: USA
;; ZIP: 19103-2307
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/399,986B
;; FILING DATE: 06-MAR-1995
;; CLASSIFICATION: 530
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Hagan, Patrick J.
;; REGISTRATION NUMBER: 27,643
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (215) 563-4100
;; TELEFAX: (215) 563-4044
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 2182 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: not relevant
;; MOLECULE TYPE: DNA (genomic)
;; HYPOTHETICAL: NO
;; ANTI-SENSE: NO
;; US-08-399-986B-1

Query Match 69.1%; Score 15.2; DB 1; Length 2182;
Best Local Similarity 85.0%; Pred. No. 62;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 ttggatcacagaatttgata 22
||||| || ||||| |||||
Db 1996 TTGGGCTCTCAAGATTGATA 1977

Search completed: March 27, 2001, 08:19:19
Job time: 5146 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 27, 2001, 08:19:19 ; Search time 132.3 Seconds
(without alignments)
8.527 Million cell updates/sec

Title: US-09-380-826A-4
Perfect score: 7 tgttggg 7
Sequence: 1 tgttggg 7
Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 280836 seqs, 80580151 residues

Total number of hits satisfying chosen parameters: 561672

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents, NA.*

- 1: /cgn2_6/ptodata/2/ina/5A-COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B-COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6-COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/PCTUS-COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	7	100.0	10	2	US-08-631-751A-13
C 2	7	100.0	14	1	US-08-303-004-19
C 3	7	100.0	14	3	US-08-765-340-154
C 4	7	100.0	15	1	US-08-182-968A-201
C 5	7	100.0	15	1	US-08-182-968A-246
C 6	7	100.0	15	1	US-08-363-240A-614
C 7	7	100.0	15	1	US-08-363-240A-615
C 8	7	100.0	15	1	US-08-363-240A-616
C 9	7	100.0	15	2	US-08-774-306A-201
C 10	7	100.0	15	2	US-08-774-306A-246
C 11	7	100.0	15	2	US-08-585-684B-1366
C 12	7	100.0	15	2	US-08-585-684B-1367
C 13	7	100.0	15	2	US-08-585-684B-1729
C 14	7	100.0	15	2	US-08-585-684B-1730
C 15	7	100.0	15	2	US-08-585-684B-1731
C 16	7	100.0	15	2	US-08-585-684B-1732
C 17	7	100.0	15	2	US-08-585-684B-1768
C 18	7	100.0	15	2	US-08-585-684B-1769
C 19	7	100.0	15	2	US-08-585-684B-1770
C 20	7	100.0	15	2	US-08-585-684B-2139
C 21	7	100.0	15	3	US-09-064-156A-201
C 22	7	100.0	15	3	US-09-064-156A-246
C 23	7	100.0	16	1	US-08-435-350-32
C 24	7	100.0	16	1	US-08-166-664-15
C 25	7	100.0	16	2	US-08-469-461-15
C 26	7	100.0	16	3	US-07-890-609-15
C 27	7	100.0	17	1	US-08-072-063-4
C 28	7	100.0	17	1	US-08-119-773-23

C 29	7	100.0	17	1	US-08-281-940-25	Sequence 25, Appl
C 30	7	100.0	17	1	US-08-390-850-433	Sequence 433, App
C 31	7	100.0	17	1	US-08-064-693-4	Sequence 4, Appli
C 32	7	100.0	17	1	US-08-373-124A-810	Sequence 810, App
C 33	7	100.0	17	1	US-08-441-370-2	Sequence 2, Appli
C 34	7	100.0	17	1	US-08-435-634-433	Sequence 433, App
C 35	7	100.0	17	1	US-08-433-628-810	Sequence 810, App
C 36	7	100.0	17	2	US-08-710-134-25	Sequence 25, Appl
C 37	7	100.0	17	2	US-08-485-885-25	Sequence 25, Appl
C 38	7	100.0	17	3	US-08-704-966-5	Sequence 5, Appli
C 39	7	100.0	17	3	US-08-705-438-5	Sequence 5, Appli
C 40	7	100.0	17	3	US-08-985-162-529	Sequence 529, App
C 41	7	100.0	17	3	US-08-985-162-530	Sequence 530, App
C 42	7	100.0	17	3	US-08-985-162-531	Sequence 531, App
C 43	7	100.0	17	4	PCT-US93-04754-4	Sequence 4, Appli
C 44	7	100.0	18	1	US-08-361-920-47	Sequence 47, Appl
C 45	7	100.0	18	1	US-08-479-939-47	Sequence 47, Appl

ALIGNMENTS

RESULT 1
US-08-631-751A-13/C
; Sequence 13, Application US/08631751A
; Patent No. 5843767
; GENERAL INFORMATION:
; APPLICANT: Beattie, Kenneth L.
; TITLE OF INVENTION: Microfabricated, Flowthrough Porous
; TITLE OF INVENTION: Apparatus for Discrete Detection of Binding Reactions
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Vinson & Elkins
; STREET: 1455 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20004-1008
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/631,751A
; FILING DATE: 11-April-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sanzo, Michael A.
; REGISTRATION NUMBER: 36,912
; REFERENCE/DOCKET NUMBER: HARC0001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)639-6500
; TELEFAX: (202)639-6604
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
US-08-631-751A-13

Query Match 100.0%; Score 7; DB 2; Length 10;
Best Local Similarity 100.0%; Pred. No. 5.1e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgttggg 7
Db 7 TGTGTGA 1

RESULT 2

```
US-08-303-004-19/C
; Sequence 19, Application US/08303004
; Patent No. 5556955
; GENERAL INFORMATION:
; APPLICANT: Vergnaud, Gilles
; TITLE OF INVENTION: Process for Detection of New Polymor-
; TITLE OF INVENTION: phic Loci in an ADN Sequence, Nucleotide Sequences Forming
; TITLE OF INVENTION: Hybridisation Probes and Their Biological Applications
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Oliff & Berridge
; STREET: P.O. Box 19928
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; APPLICATION NUMBER: US/08/303,004
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/931,311B
; FILING DATE: 19920818
; ATTORNEY/AGENT INFORMATION:
; NAME: Berridge, William P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: WPB 28264
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6400
; TELEFAX: (703) 836-2787
; TELEX: 90-1799 PTO ALEX
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; US-08-303-004-19

Query Match 100.0%; Score 7; DB 1; Length 14;
Best Local Similarity 100.0%; Pred. No. 5.2e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgttggg 7
Db 12 TGTGGA 6

RESULT 3
US-08-765-340-154
; Sequence 154, Application US/08765340
; Patent No. 6150092
; GENERAL INFORMATION:
; APPLICANT: UCHIDA, K.,
; APPLICANT: UCHIDA, T.,
; APPLICANT: TANAKA, Y.,
; APPLICANT: MATSUDA, Y.,
; APPLICANT: KONDO, S.
; TITLE OF INVENTION: AN ANTISENSE NUCLEIC ACID
; NUMBER OF SEQUENCES: 185
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
; STREET: 345 PARK AVENUE
```

```
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version
; SOFTWARE: #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/765,340
; FILING DATE: 23-DEC-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 145146/94
; FILING DATE: 27-JUN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 311130/94
; FILING DATE: 21-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: SERUNIAN, LESLIE
; REGISTRATION NUMBER: 35,353
; REFERENCE/DOCKET NUMBER: 1452-4005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 154:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "synthetic DNA"
; US-08-765-340-154

Query Match 100.0%; Score 7; DB 3; Length 14;
Best Local Similarity 100.0%; Pred. No. 5.2e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgttggg 7
Db 7 TGTGGA 13

RESULT 4
US-08-182-968A-201
; Sequence 201, Application US/08182968A
; Patent No. 5610054
; GENERAL INFORMATION:
; APPLICANT: Draper, Kenneth G.
; TITLE OF INVENTION: METHOD AND REAGENT FOR
; TITLE OF INVENTION: INHIBITING HEPATITIS C
; TITLE OF INVENTION: VIRUS REPLICATION
; NUMBER OF SEQUENCES: 497
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/182,968A
; FILING DATE: 13-JANUARY-1994
```


PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/882,888
FILING DATE: 14-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 205/277
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 201:
SEQUENCE CHARACTERISTICS:
LENGTH: 15
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-182-968A-201

Query Match 100.0%; Score 7; DB 1; Length 15;
Best Local Similarity 57.1%; Pred. No. 5.2e+03;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgttga 7
Db 5 UGUUGGA 11

RESULT 5
US-08-182-968A-246/c
Sequence 246, Application US/08182968A
Patent No. 5610054
GENERAL INFORMATION:
APPLICANT: Draper, Kenneth G.
TITLE OF INVENTION: METHOD AND REAGENT FOR
TITLE OF INVENTION: INHIBITING HEPATITIS C
TITLE OF INVENTION: VIRUS REPLICATION
NUMBER OF SEQUENCES: 497
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
CITY: Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071-2066
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/182,968A
FILING DATE: 13-JANUARY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/882,888
FILING DATE: 14-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 205/277
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 246:
SEQUENCE CHARACTERISTICS:
LENGTH: 15
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

US-08-182-968A-246

Query Match 100.0%; Score 7; DB 1; Length 15;
Best Local Similarity 100.0%; Pred. No. 5.2e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgttga 7
Db 13 TGTGTGA 7

RESULT 6
US-08-363-240A-614/c
Sequence 614, Application US/08363240A
Patent No. 5705388
GENERAL INFORMATION:
APPLICANT: Couture, Larry
APPLICANT: McSwiggen, James
APPLICANT: Bisgaler, Charles
APPLICANT: Pape, Michael
TITLE OF INVENTION: METHOD AND REAGENT FOR
TITLE OF INVENTION: PREVENTION, INHIBITION OF
TITLE OF INVENTION: PROGRESSION AND REGRESSION
TITLE OF INVENTION: OF VASCULAR DISEASES
NUMBER OF SEQUENCES: 1243
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
CITY: Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: Storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/363,240A
FILING DATE: December 23, 1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 210/096
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 614:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-363-240A-614

Query Match 100.0%; Score 7; DB 1; Length 15;
Best Local Similarity 100.0%; Pred. No. 5.2e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgttga 7
Db 15 TGTGTGA 9

RESULT 7

US-08-363-240A-615/C
 ; Sequence 615, Application US/08363240A
 ; Patent No. 5705388
 ; GENERAL INFORMATION:
 ; APPLICANT: Couture, Larry
 ; APPLICANT: McSwiggen, James
 ; APPLICANT: Bisgaier, Charles
 ; APPLICANT: Pape, Michael
 ; TITLE OF INVENTION: METHOD AND REAGENT FOR
 ; TITLE OF INVENTION: PREVENTION, INHIBITION OF
 ; TITLE OF INVENTION: PROGRESSION AND REGRESSION
 ; TITLE OF INVENTION: OF VASCULAR DISEASES
 ; NUMBER OF SEQUENCES: 1243
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Lyon & Lyon
 ; STREET: 633 West Fifth Street
 ; CITY: Los Angeles
 ; STATE: California
 ; COUNTRY: U.S.A.
 ; ZIP: 90071
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
 ; MEDIUM TYPE: storage
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: IBM P.C. DOS 5.0
 ; SOFTWARE: Word Perfect 5.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/363,240A
 ; FILING DATE: December 23, 1994
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Warburg, Richard
 ; REGISTRATION NUMBER: 32,327
 ; REFERENCE/DOCKET NUMBER: 210/096
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (213) 489-1600
 ; TELEFAX: (213) 955-0440
 ; TELEX: 67-3510
 ; INFORMATION FOR SEQ ID NO: 615:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 15 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-08-363-240A-615

Query Match 100.0%; Score 7; DB 1; Length 15;
 Best Local Similarity 100.0%; Pred. No. 5.2e+03;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 tgttggg 7
 Db 15 TGTGGGA 9
 RESULT 8
 US-08-363-240A-616/C
 ; Sequence 616, Application US/08363240A
 ; Patent No. 5705388
 ; GENERAL INFORMATION:
 ; APPLICANT: Couture, Larry
 ; APPLICANT: McSwiggen, James
 ; APPLICANT: Bisgaier, Charles
 ; APPLICANT: Pape, Michael
 ; TITLE OF INVENTION: METHOD AND REAGENT FOR
 ; TITLE OF INVENTION: PREVENTION, INHIBITION OF
 ; TITLE OF INVENTION: PROGRESSION AND REGRESSION
 ; TITLE OF INVENTION: OF VASCULAR DISEASES
 ; NUMBER OF SEQUENCES: 1243

US-08-363-240A-615/C
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Lyon & Lyon
 ; STREET: 633 West Fifth Street
 ; CITY: Los Angeles
 ; STATE: California
 ; COUNTRY: U.S.A.
 ; ZIP: 90071
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
 ; MEDIUM TYPE: storage
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: IBM P.C. DOS 5.0
 ; SOFTWARE: Word Perfect 5.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/363,240A
 ; FILING DATE: December 23, 1994
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Warburg, Richard
 ; REGISTRATION NUMBER: 32,327
 ; REFERENCE/DOCKET NUMBER: 210/096
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (213) 489-1600
 ; TELEFAX: (213) 955-0440
 ; TELEX: 67-3510
 ; INFORMATION FOR SEQ ID NO: 616:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 15 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-08-363-240A-616
 Query Match 100.0%; Score 7; DB 1; Length 15;
 Best Local Similarity 100.0%; Pred. No. 5.2e+03;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 tgttggg 7
 Db 13 TGTGGGA 7
 RESULT 9
 US-08-774-306A-201
 ; Sequence 201, Application US/08774306A
 ; Patent No. 5869253
 ; GENERAL INFORMATION:
 ; APPLICANT: Draper, Kenneth G.
 ; TITLE OF INVENTION: METHOD AND REAGENT FOR
 ; TITLE OF INVENTION: INHIBITING HEPATITIS C
 ; TITLE OF INVENTION: VIRUS REPLICATION
 ; NUMBER OF SEQUENCES: 497
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Lyon & Lyon
 ; STREET: 633 West Fifth Street
 ; CITY: Los Angeles
 ; STATE: California
 ; COUNTRY: U.S.A.
 ; ZIP: 90071-2066
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
 ; MEDIUM TYPE: storage
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: IBM P.C. DOS 5.0
 ; SOFTWARE: Word Perfect 5.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/774,306A
 ; FILING DATE: December 26, 1996

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/182,968
; FILING DATE: January 13, 1994
; APPLICATION NUMBER: 07/882,888
; FILING DATE: May 14, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 223/227
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 201:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-774-306A-201

```

```

Query Match 100.0%; Score 7; DB 2; Length 15;
Best Local Similarity 57.1%; Pred. No. 5.2e+03;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 tgttga 7
   :||:|
Db 5 UGUUGA 11

```

```

RESULT 10
US-08-774-306A-246/c
; Sequence 246, Application US/08774306A
; Patent No. 5869253
; GENERAL INFORMATION:
; APPLICANT: Draper, Kenneth G.
; TITLE OF INVENTION: METHOD AND REAGENT FOR
; TITLE OF INVENTION: INHIBITING HEPATITIS C
; NUMBER OF SEQUENCES: 497
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/774,306A
; FILING DATE: December 26, 1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/182,968
; FILING DATE: January 13, 1994
; APPLICATION NUMBER: 07/882,888
; FILING DATE: May 14, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 223/227
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 246:
; SEQUENCE CHARACTERISTICS:

```

```

; LENGTH: 15
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-774-306A-246

Query Match 100.0%; Score 7; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 5.2e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgttga 7
   :||:|
Db 13 TGTGGA 7

RESULT 11
US-08-585-684B-1366/c
; Sequence 1366, Application US/08585684B
; Patent No. 5877021
; GENERAL INFORMATION:
; APPLICANT: Stinchcomb, Daniel T.
; APPLICANT: Jarvis, Thale
; APPLICANT: McSwiggan, James
; TITLE OF INVENTION: METHOD AND REAGENT FOR THE
; TITLE OF INVENTION: INDUCTION OF GRAFT TOLERANCE
; TITLE OF INVENTION: AND REVERSAL OF IMMUNE RESPONSES
; NUMBER OF SEQUENCES: 2751
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: FastSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/585,684B
; FILING DATE: January 16, 1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/000,951
; FILING DATE: July 7, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 218/078
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 1366:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-585-684B-1366

```

```

Query Match 100.0%; Score 7; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 5.2e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 tgttga 7
   :||:|
Db 14 TGTGGA 8

```

RESULT 12
US-08-585-684B-1367/C
; Sequence 1367, Application US/08585684B
; Patent No. 5877021
; GENERAL INFORMATION:
; APPLICANT: Stinchcomb, Daniel T.
; APPLICANT: Jarvis, Thale
; APPLICANT: McSwiggen, James
; TITLE OF INVENTION: METHOD AND REAGENT FOR THE
; TITLE OF INVENTION: INDUCTION OF GRAFT TOLERANCE
; TITLE OF INVENTION: AND REVERSAL OF IMMUNE RESPONSES
; NUMBER OF SEQUENCES: 2751
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: FastSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/585,684B
; FILING DATE: January 16, 1996
; PRIOR APPLICATION NUMBER: 60/000,951
; FILING DATE: July 7, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 218/078
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 1367:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-585-684B-1367

Query Match 100.0%; Score 7; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 5.2e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgttggga 7
| | | | | | |
Db 13 TGTGGA 7

RESULT 13
US-08-585-684B-1729/C
; Sequence 1729, Application US/08585684B
; Patent No. 5877021
; GENERAL INFORMATION:
; APPLICANT: Stinchcomb, Daniel T.
; APPLICANT: Jarvis, Thale
; APPLICANT: McSwiggen, James
; TITLE OF INVENTION: METHOD AND REAGENT FOR THE
; TITLE OF INVENTION: INDUCTION OF GRAFT TOLERANCE
; TITLE OF INVENTION: AND REVERSAL OF IMMUNE RESPONSES
; NUMBER OF SEQUENCES: 2751
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon

STREET: 633 West Fifth Street
STREET: Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: FastSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/585,684B
FILING DATE: January 16, 1996
PRIOR APPLICATION NUMBER: 60/000,951
FILING DATE: July 7, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 218/078
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 1729:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-585-684B-1729

Query Match 100.0%; Score 7; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 5.2e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgttggga 7
| | | | | | |
Db 13 TGTGGA 7

RESULT 14
US-08-585-684B-1730/C
; Sequence 1730, Application US/08585684B
; Patent No. 5877021
; GENERAL INFORMATION:
; APPLICANT: Stinchcomb, Daniel T.
; APPLICANT: Jarvis, Thale
; APPLICANT: McSwiggen, James
; TITLE OF INVENTION: METHOD AND REAGENT FOR THE
; TITLE OF INVENTION: INDUCTION OF GRAFT TOLERANCE
; TITLE OF INVENTION: AND REVERSAL OF IMMUNE RESPONSES
; NUMBER OF SEQUENCES: 2751
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: FastSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/585,684B
FILING DATE: January 16, 1996

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/000,951
; FILING DATE: July 7, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 218/078
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 1730:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-585-684B-1730

```

```

Query Match 100.0%; Score 7; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 5.2e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 tgttga 7
Db 13 TGTGGA 7

```

RESULT 15

```

US-08-585-684B-1731/C
; Sequence 1731, Application US/08585684B
; Patent No. 5877021
; GENERAL INFORMATION:

```

```

; APPLICANT: Stinchcomb, Daniel T.
; APPLICANT: Jarvis, Thale
; APPLICANT: McSwiggen, James
; TITLE OF INVENTION: METHOD AND REAGENT FOR THE
; TITLE OF INVENTION: INDUCTION OF GRAFT TOLERANCE
; TITLE OF INVENTION: AND REVERSAL OF IMMUNE RESPONSES
; NUMBER OF SEQUENCES: 2751
; CORRESPONDENCE ADDRESS:

```

```

; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/585,684B
; FILING DATE: January 16, 1996

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/000,951
; FILING DATE: July 7, 1995
; ATTORNEY/AGENT INFORMATION:

```

```

; NAME: Warburg, Richard
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 218/078
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510

```

```

; INFORMATION FOR SEQ ID NO: 1731:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 base pairs
; TYPE: nucleic acid

```

```

; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-585-684B-1731

```

```

Query Match 100.0%; Score 7; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 5.2e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 tgttga 7
Db 13 TGTGGA 7

```

```

Search completed: March 27, 2001, 08:19:22
Job time: 5149 sec

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 27, 2001, 08:22:59 ; Search time 207.51 Seconds
(without alignments)
12.672 Million cell updates/sec

Title: US-09-380-826A-5

Perfect score: 7

Sequence: 1 ttgata 7

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 480072 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_36:*

- 1: /cgn2_2/gcgdata/geneseq/geneseq/NA1980.DAT:*
- 2: /cgn2_2/gcgdata/geneseq/geneseq/NA1981.DAT:*
- 3: /cgn2_2/gcgdata/geneseq/geneseq/NA1982.DAT:*
- 4: /cgn2_2/gcgdata/geneseq/geneseq/NA1983.DAT:*
- 5: /cgn2_2/gcgdata/geneseq/geneseq/NA1984.DAT:*
- 6: /cgn2_2/gcgdata/geneseq/geneseq/NA1985.DAT:*
- 7: /cgn2_2/gcgdata/geneseq/geneseq/NA1986.DAT:*
- 8: /cgn2_2/gcgdata/geneseq/geneseq/NA1987.DAT:*
- 9: /cgn2_2/gcgdata/geneseq/geneseq/NA1988.DAT:*
- 10: /cgn2_2/gcgdata/geneseq/geneseq/NA1989.DAT:*
- 11: /cgn2_2/gcgdata/geneseq/geneseq/NA1990.DAT:*
- 12: /cgn2_2/gcgdata/geneseq/geneseq/NA1991.DAT:*
- 13: /cgn2_2/gcgdata/geneseq/geneseq/NA1992.DAT:*
- 14: /cgn2_2/gcgdata/geneseq/geneseq/NA1993.DAT:*
- 15: /cgn2_2/gcgdata/geneseq/geneseq/NA1994.DAT:*
- 16: /cgn2_2/gcgdata/geneseq/geneseq/NA1995.DAT:*
- 17: /cgn2_2/gcgdata/geneseq/geneseq/NA1996.DAT:*
- 18: /cgn2_2/gcgdata/geneseq/geneseq/NA1997.DAT:*
- 19: /cgn2_2/gcgdata/geneseq/geneseq/NA1998.DAT:*
- 20: /cgn2_2/gcgdata/geneseq/geneseq/NA1999.DAT:*
- 21: /cgn2_2/gcgdata/geneseq/geneseq/NA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7	100.0	7	19	V58899
2	7	100.0	10	21	278577
3	7	100.0	10	21	280864
4	7	100.0	10	21	282567
5	7	100.0	10	21	282768
6	7	100.0	10	21	283117
7	7	100.0	13	21	297719
8	7	100.0	14	19	V49155
9	7	100.0	14	21	297691
10	7	100.0	15	14	052243
11	7	100.0	15	16	T56960
12	7	100.0	15	16	T56962

13	7	100.0	15	16	T56964	RSV IC hammerhead
c 14	7	100.0	15	16	T57424	RSV N hammerhead r
c 15	7	100.0	15	16	T57425	RSV N hammerhead r
16	7	100.0	15	16	T54843	Mouse re1A hammerh
17	7	100.0	15	16	T54845	Mouse re1A hammerh
18	7	100.0	15	16	T54841	Mouse re1A hammerh
c 19	7	100.0	15	17	X56632	Human CD40 hammerh
c 20	7	100.0	15	17	X56633	Human CD40 hammerh
c 21	7	100.0	15	17	X56634	Human CD40 hammerh
c 22	7	100.0	15	17	X56635	Human CD40 hammerh
c 23	7	100.0	15	17	X56825	Mouse CD40 hammerh
c 24	7	100.0	15	17	X56826	Mouse CD40 hammerh
c 25	7	100.0	15	17	X56827	Mouse CD40 hammerh
26	7	100.0	15	17	X64619	Human B7-1 hammerh
27	7	100.0	15	17	X64620	Human B7-1 hammerh
28	7	100.0	15	17	X64621	Human B7-1 hammerh
29	7	100.0	15	17	X64622	Human B7-1 hammerh
c 30	7	100.0	15	17	X75694	Human flt-1 and KD
c 31	7	100.0	15	20	X76904	H2-1 Pagl gene dir
c 32	7	100.0	15	20	X31695	Transcript tag seq
c 33	7	100.0	15	20	X31205	Tag sequence of a
c 34	7	100.0	15	21	297690	HIV-1 protease gen
c 35	7	100.0	15	21	297727	HIV-1 protease gen
c 36	7	100.0	16	12	Q11395	Probe COD 1111 Spe
c 37	7	100.0	16	14	Q52019	B-cell mRNA ribozy
c 38	7	100.0	16	16	Q74138	Basic fibroblast g
c 39	7	100.0	16	21	297686	HIV-1 protease gen
c 40	7	100.0	16	21	297687	HIV-1 protease gen
c 41	7	100.0	16	21	297689	HIV-1 protease gen
c 42	7	100.0	16	21	297702	HIV-1 protease gen
c 43	7	100.0	16	21	297704	HIV-1 protease gen
c 44	7	100.0	16	21	297721	HIV-1 protease gen
c 45	7	100.0	16	21	297722	HIV-1 protease gen

ALIGNMENTS

RESULT 1	
V58899	ID V58899 standard; DNA; 7 BP.
XX	
AC V58899;	
XX	
DT 20-JAN-1999 (first entry)	
XX	
DE Leptosira nucleotide sequence.	
XX	
KW Infection; pathogenic Leptosira; protective immunity; therapy;	
diagnosis; ss.	
XX	
OS Leptosira sp.	
XX	
PN WO9840099-A1.	
XX	
PD 17-SEP-1998.	
XX	
PF 06-MAR-1998; 98WO-AU00145.	
XX	
PR 07-MAR-1997; 97AU-0005494.	
XX	
PA (AGRI-) AGRIC VICTORIA SERVICES PTY LTD.	
PA (PIGR-) PIG RES & DEV CORP.	
PI	Chappel RJ;
XX	
DR WPI; 1998-520791/44.	
XX	
PT New isolated pathogenic Leptosira bacterium - useful for, e.g	
PT developing products for conferring protective immunity, and for	
PT prophylactic or therapeutic treatment	
PS Claim 15; Page 72; 94pp; English.	

XX This sequence represents a Leptospira DNA sequence isolated from the
 CC pathogenic Leptospira (LS) bacterium of the invention. The bacterium
 CC belongs to serogroup Hurstbridge or serovar Hurstbridge or the species
 CC L. fainei. The LS bacteria can be used for conferring protective
 CC immunity against pathogenic LS bacteria in humans or animals. The
 CC bacteria can also be used for prophylactic or therapeutic treatment of LS
 CC infections. The DNAs and antibodies may also be used for detection and
 CC diagnosis of past or present LS infection.
 XX
 SQ Sequence 7 BP; 2 A; 0 C; 1 G; 4 T; 0 other;

Query Match 100.0%; Score 7; DB 19; Length 7;
 Best Local Similarity 100.0%; Pred. No. 5.2e+07;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ttgtgata 7
 |||||
 Db 1 ttgtgata 7

RESULT 2
 278577/C
 ID 278577 standard; DNA; 10 BP.

XX 278577;

DT 10-APR-2000 (first entry)

DE Human dendritic cell SAGE tag, SEQ ID NO:1005.

XX SAGE tag: serial analysis of gene expression; antigen-presenting cell;
 KW APC; monocyte-derived dendritic cell; differential gene expression;
 KW immunostimulatory cofactor; costimulatory factor; CTL;
 KW cytotoxic T-lymphocyte; tumour antigen; immunotherapy; anticancer; ss.

XX Homo sapiens.

XX WO9965924-A2.

PN 23-DEC-1999.

PD 18-JUN-1999; 99WO-US13800.

XX 19-JUN-1998; 98US-0089833.
 XX 19-JUN-1998; 98US-0089844.
 XX 19-JUN-1998; 98US-0089853.
 XX 19-JUN-1998; 98US-0089878.
 XX 19-JUN-1998; 98US-0089991.
 XX 19-JUN-1998; 98US-0089992.
 XX 19-JUN-1998; 98US-0089993.
 XX 19-JUN-1998; 98US-0089994.
 XX 19-JUN-1998; 98US-0089997.
 XX 19-JUN-1998; 98US-0089999.
 XX 19-JUN-1998; 98US-0090000.
 XX 19-JUN-1998; 98US-0090035.
 XX 19-JUN-1998; 98US-0090036.
 XX 19-JUN-1998; 98US-0090039.
 XX 19-JUN-1998; 98US-0090040.
 XX 19-JUN-1998; 98US-0090041.
 XX 19-JUN-1998; 98US-0090042.
 XX 19-JUN-1998; 98US-0090043.
 XX 19-JUN-1998; 98US-0090044.
 XX 19-JUN-1998; 98US-0090045.
 XX 19-JUN-1998; 98US-0090047.
 XX 19-JUN-1998; 98US-0090048.
 XX 19-JUN-1998; 98US-0090072.
 XX 19-JUN-1998; 98US-0090076.
 XX 19-JUN-1998; 98US-0090077.
 XX 19-JUN-1998; 98US-0090078.
 XX 19-JUN-1998; 98US-0090079.
 XX 19-JUN-1998; 98US-0090080.

PR 08-DEC-1998; 98US-0111715.

XX (GENZ) GENZYME CORP.
 PA (ROBE/) ROBERTS B L.
 PA (SHAN/) SHANKARA S.

XX Roberts BL, Shankara S;

XX WPI; 2000-106077/09.

XX Isolated polynucleotides differentially expressed in antigen-presenting
 PT cells, useful in gene vaccines against cancer -

XX Claim 1; Page 94; 130pp; English.

XX Sequences 277573-279709 represent SAGE (serial analysis of gene
 CC expression) tags used to identify mRNA transcripts encoding
 CC immunostimulatory cofactor proteins which are preferentially or
 CC differentially expressed in monocyte-derived dendritic cells compared
 CC with monocytes. Some of the transcripts correspond to known genes or
 CC ESTs (expressed sequence tags) which were previously unknown to be
 CC preferentially or differentially expressed in dendritic cells, while
 CC other transcripts correspond to novel genes. Antigen-presenting cell
 CC (APC)-associated costimulatory factors play an important role in the
 CC activation of the cytotoxic immune response, particularly against tumour
 CC cells. Tumour antigen presentation via the MHC (major histocompatibility
 CC complex) and subsequent recognition by T-cell receptors is alone
 CC insufficient to activate a robust cytotoxic immune response that can lyse
 CC the tumour cells. Immunostimulatory cofactors also being required for
 CC efficient activation of cytotoxic T-lymphocytes (CTLs). Nucleic acid
 CC sequences identified using the SAGE tags have several potential uses.
 CC They may be used in vaccines to induce an immune response, particularly
 CC against a tumour antigen; to modulate the genotype of an APC; to screen
 CC for agents that modulate expression of differentially expressed genes in
 CC an APC; and as hybridisation probes/amplification primers for the
 CC diagnosis, prognosis and monitoring of diseases related to abnormal
 CC expression of these genes. Detection of the dendritic cell differentially
 CC expressed genes, or of their encoded proteins, can be used to identify
 CC cells as belonging to the monocyte lineage. Cells containing these genes
 CC can be used in active immunotherapy (or to stimulate production of a
 CC population of antigen-specific effector cells) and vectors containing
 CC them are used in gene therapy. Co-administration of tumour antigens and
 CC APC-associated costimulatory factors ensures adequate antigen
 CC presentation to endogenous APCs and upregulates the APCs for the
 CC presentation of co-stimulatory signals, migration to T cell-rich sites,
 CC secretion of T cell growth factors and secretion of chemokines for
 CC recruitment of immune effector cells.

XX Sequence 10 BP; 6 A; 1 C; 1 G; 2 T; 0 other;

Query Match 100.0%; Score 7; DB 21; Length 10;

Best Local Similarity 100.0%; Pred. No. 1.9e+04;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ttgtgata 7

Db 9 tttgata 3

RESULT 3

Z80864

ID Z80864 standard; DNA; 10 BP.

XX Z80864;

XX 07-APR-2000 (first entry)

DE Metastatic breast tumour cell upregulated transcript tag #98.

XX Human; metastatic breast tumour tissue; breast cancer; tag; primer;
 KW non-metastatic breast tumour tissue; gene therapy; anticancer;
 KW antimetastatic; vaccine; diagnosis; ss.

XX Homo sapiens.
 XX WO965928-A2.
 XX 23-DEC-1999.
 XX 18-JUN-1999; 99WO-US13647.
 XX 19-JUN-1998; 98US-0089853.
 XX 19-JUN-1998; 98US-0089997.
 XX 19-JUN-1998; 98US-0090039.
 XX 19-JUN-1998; 98US-0090040.
 XX 19-JUN-1998; 98US-0090041.
 XX (GENZ) GENZYME CORP.
 XX (ROBE/) ROBERTS B L.
 XX (SHAN/) SHANKARA S.
 XX Roberts BL, Shankara S;
 XX WPI; 2000-106079/09.
 XX Isolated polynucleotides differentially expressed between metastatic
 and non-metastatic breast cancer cells, useful for diagnosis,
 prevention and treatment of cancer -
 Claim 1; Page 60; 219pp; English.
 XX 280767 to 283941 represent tags corresponding to distinct transcripts
 that are preferentially transcribed in the metastatic breast tumour
 tissue (i.e. are upregulated in metastatic breast tumour cells). 283942
 to 286677 represent tags corresponding to distinct transcripts that are
 preferentially transcribed in the primary or non-metastatic breast tumour
 tissue (i.e. are downregulated in metastatic breast tumour cells).
 These transcripts can be used for diagnosis, prognosis, monitoring and
 treatment of breast cancer, particularly where metastatic. Diagnosis is
 by standard immunoassays or hybridisation/amplification reactions.
 Compounds that modulate expression of the transcripts are potentially
 useful for treatment of (metastatic) breast cancer, while promoters from
 the transcripts are used to direct expression, in selected cell types,
 of e.g. therapeutic genes (also ribozymes or antisense sequences),
 particularly an antigen-encoding sequence for use in gene or cell-based
 vaccines. Polypeptides encoded by the transcripts are also useful in
 vaccines; for diagnosing breast cancer and for raising specific
 antibodies (Ab). Ab are used to detect the polypeptides or as therapeutic
 agents. Host cells that produce the polypeptides can be used to expand
 and isolate populations of educated, antigen-specific immune effector
 cells, e.g. cytotoxic T lymphocytes, and these used for adoptive
 immunotherapy.
 XX Sequence 10 BP; 3 A; 0 C; 1 G; 6 T; 0 other;
 Query Match 100.0%; Score 7; DB 21; Length 10;
 Best Local Similarity 100.0%; Pred No. 1.9e+04;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ttgtata 7
 |||||
 Db 3 ttgtata 9
 RESULT 4
 282567
 ID 282567 standard; DNA; 10 BP.
 XX
 AC 282567;
 XX
 DT 07-APR-2000 (first entry)
 XX
 DE Metastatic breast tumour cell upregulated transcript tag #1801.
 XX

KW Human; metastatic breast tumour tissue; breast cancer; tag; primer;
 KW non-metastatic breast tumour tissue; gene therapy; anticancer;
 KW antimetastatic; vaccine; diagnosis; ss.
 XX Homo sapiens.
 XX WO965928-A2.
 XX 23-DEC-1999.
 XX 18-JUN-1999; 99WO-US13647.
 XX 19-JUN-1998; 98US-0089853.
 XX 19-JUN-1998; 98US-0089997.
 XX 19-JUN-1998; 98US-0090039.
 XX 19-JUN-1998; 98US-0090040.
 XX 19-JUN-1998; 98US-0090041.
 XX (GENZ) GENZYME CORP.
 XX (ROBE/) ROBERTS B L.
 XX (SHAN/) SHANKARA S.
 XX Roberts BL, Shankara S;
 XX WPI; 2000-106079/09.
 XX Isolated polynucleotides differentially expressed between metastatic
 and non-metastatic breast cancer cells, useful for diagnosis,
 prevention and treatment of cancer -
 Claim 1; Page 107; 219pp; English.
 XX 280767 to 283941 represent tags corresponding to distinct transcripts
 that are preferentially transcribed in the metastatic breast tumour
 tissue (i.e. are upregulated in metastatic breast tumour cells). 283942
 to 286677 represent tags corresponding to distinct transcripts that are
 preferentially transcribed in the primary or non-metastatic breast tumour
 tissue (i.e. are downregulated in metastatic breast tumour cells).
 These transcripts can be used for diagnosis, prognosis, monitoring and
 treatment of breast cancer, particularly where metastatic. Diagnosis is
 by standard immunoassays or hybridisation/amplification reactions.
 Compounds that modulate expression of the transcripts are potentially
 useful for treatment of (metastatic) breast cancer, while promoters from
 the transcripts are used to direct expression, in selected cell types,
 of e.g. therapeutic genes (also ribozymes or antisense sequences),
 particularly an antigen-encoding sequence for use in gene or cell-based
 vaccines. Polypeptides encoded by the transcripts are also useful in
 vaccines; for diagnosing breast cancer and for raising specific
 antibodies (Ab). Ab are used to detect the polypeptides or as therapeutic
 agents. Host cells that produce the polypeptides can be used to expand
 and isolate populations of educated, antigen-specific immune effector
 cells, e.g. cytotoxic T lymphocytes, and these used for adoptive
 immunotherapy.
 XX Sequence 10 BP; 3 A; 0 C; 2 G; 5 T; 0 other;
 Query Match 100.0%; Score 7; DB 21; Length 10;
 Best Local Similarity 100.0%; Pred No. 1.9e+04;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ttgtata 7
 |||||
 Db 1 ttgtata 7
 RESULT 5
 282768
 ID 282768 standard; DNA; 10 BP.
 XX
 AC 282768;
 XX
 DT 07-APR-2000 (first entry)

XX DE Metastatic breast tumour cell upregulated transcript tag #2002.
 XX XX
 KW Human; metastatic breast tumour tissue; breast cancer; tag; primer;
 KW non-metastatic breast tumour tissue; gene therapy; anticancer;
 KW antimetastatic; vaccine; diagnosis; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO9965928-A2.
 XX
 PD 23-DEC-1999.
 XX
 PE 18-JUN-1999; 99WO-US13647.
 XX
 PF 19-JUN-1998; 98US-0089853.
 PR 19-JUN-1998; 98US-0089997.
 PR 19-JUN-1998; 98US-0090039.
 PR 19-JUN-1998; 98US-0090040.
 PR 19-JUN-1998; 98US-0090041.
 XX
 PA (GENZ) GENZYME CORP.
 PA (ROBE/) ROBERTS B L.
 PA (SHAN/) SHANKARA S.
 XX
 PI Roberts BL, Shankara S;
 XX
 DR WPI; 2000-106079/09.
 XX
 PT Isolated polynucleotides differentially expressed between metastatic
 PT and non-metastatic breast cancer cells, useful for diagnosis,
 PT prevention and treatment of cancer -
 XX
 PS Claim 1: Page 113; 219pp; English.
 XX
 CC 280767 to 283941 represent tags corresponding to distinct transcripts
 CC that are preferentially transcribed in the metastatic breast tumour
 CC tissue (i.e. are upregulated in metastatic breast tumour cells). 283942
 CC to 286677 represent tags corresponding to distinct transcripts that are
 CC preferentially transcribed in the primary or non-metastatic breast tumour
 CC tissue (i.e. are downregulated in metastatic breast tumour cells).
 CC These transcripts can be used for diagnosis, prognosis, monitoring and
 CC treatment of breast cancer, particularly where metastatic. Diagnosis is
 CC by standard immunoassays or hybridisation/amplification reactions.
 CC Compounds that modulate expression of the transcripts are potentially
 CC useful for treatment of (metastatic) breast cancer, while promoters from
 CC of e.g. therapeutic genes (also ribozymes or antisense sequences),
 CC particularly an antigen-encoding sequence for use in gene or cell-based
 CC vaccines. Polypeptides encoded by the transcripts are also useful in
 CC antibodies (Ab). Ab are used to detect the polypeptides or as therapeutic
 CC agents. Host cells that produce the polypeptides can be used to expand
 CC and isolate populations of educated, antigen-specific immune effector
 CC cells, e.g. cytotoxic T lymphocytes, and these used for adoptive
 CC immunotherapy.
 XX
 SQ Sequence 10 BP; 4 A; 0 C; 2 G; 4 T; 0 other;

Query Match 100.0%; Score 7; DB 21; Length 10;
 Best Local Similarity 100.0%; Pred. No. 1.9e+04;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ttgtgata 7
 Db 2 ttgtgata 8
 |||||

RESULT 6
 283117
 ID 283117 standard; DNA; 10 BP.
 XX

AC 283117;
 XX
 DT 07-APR-2000 (first entry)
 XX
 DE Metastatic breast tumour cell upregulated transcript tag #2351.
 XX
 KW Human; metastatic breast tumour tissue; breast cancer; tag; primer;
 KW non-metastatic breast tumour tissue; gene therapy; anticancer;
 KW antimetastatic; vaccine; diagnosis; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO9965928-A2.
 XX
 PD 23-DEC-1999.
 XX
 PE 18-JUN-1999; 99WO-US13647.
 XX
 PF 19-JUN-1998; 98US-0089853.
 PR 19-JUN-1998; 98US-0089997.
 PR 19-JUN-1998; 98US-0090039.
 PR 19-JUN-1998; 98US-0090040.
 PR 19-JUN-1998; 98US-0090041.
 XX
 PA (GENZ) GENZYME CORP.
 PA (ROBE/) ROBERTS B L.
 PA (SHAN/) SHANKARA S.
 XX
 PI Roberts BL, Shankara S;
 XX
 DR WPI; 2000-106079/09.
 XX
 PT Isolated polynucleotides differentially expressed between metastatic
 PT and non-metastatic breast cancer cells, useful for diagnosis,
 PT prevention and treatment of cancer -
 XX
 PS Claim 1; Page 122; 219pp; English.
 XX
 CC 280767 to 283941 represent tags corresponding to distinct transcripts
 CC that are preferentially transcribed in the metastatic breast tumour
 CC tissue (i.e. are upregulated in metastatic breast tumour cells). 283942
 CC to 286677 represent tags corresponding to distinct transcripts that are
 CC preferentially transcribed in the primary or non-metastatic breast tumour
 CC tissue (i.e. are downregulated in metastatic breast tumour cells).
 CC These transcripts can be used for diagnosis, prognosis, monitoring and
 CC treatment of breast cancer, particularly where metastatic. Diagnosis is
 CC by standard immunoassays or hybridisation/amplification reactions.
 CC Compounds that modulate expression of the transcripts are potentially
 CC useful for treatment of (metastatic) breast cancer, while promoters from
 CC of e.g. therapeutic genes (also ribozymes or antisense sequences),
 CC particularly an antigen-encoding sequence for use in gene or cell-based
 CC vaccines. Polypeptides encoded by the transcripts are also useful in
 CC antibodies (Ab). Ab are used to detect the polypeptides or as therapeutic
 CC agents. Host cells that produce the polypeptides can be used to expand
 CC and isolate populations of educated, antigen-specific immune effector
 CC cells, e.g. cytotoxic T lymphocytes, and these used for adoptive
 CC immunotherapy.
 XX
 SQ Sequence 10 BP; 4 A; 0 C; 1 G; 5 T; 0 other;

Query Match 100.0%; Score 7; DB 21; Length 10;
 Best Local Similarity 100.0%; Pred. No. 1.9e+04;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ttgtgata 7
 Db 1 ttgtgata 7
 |||||

RESULT 7

297719/c
ID 297719 standard; DNA; 13 BP.

XX
AC

297719;

XX
DT

26-APR-2000 (first entry)

XX
DE

HIV-1 protease gene probe SEQ ID NO:209.

XX
KW

Human immunodeficiency virus; HIV; protease; probe; detection;

XX
KW

drug selected mutation; hybridisation; genotyping; infection;

XX
KW

drug resistance; ss.

XX
OS

Human immunodeficiency virus type 1.

XX
PN

WO9967428-A2.

XX
PD

29-DEC-1999.

XX
PF

22-JUN-1999; 99WO-EP04317.

XX
PR

24-JUN-1998; 98EP-0870143.

XX
PA

(INNO-) INNOGENETICS NV.

XX
PI

Stuyver L;

XX
DR

WPI; 2000-147219/13.

XX
DT

Detection of drug-selected mutations in the HIV protease gene used to

XX
PT

treat HIV infections

XX
XX

Claim 3; Page 37; 76pp; English.

XX
PS

The present invention describes the detection of drug-selected mutations in the HIV protease gene. The method of detection allows the simultaneous characterisation of a range of codons involved in drug resistance using sets of probes optimised to function together in a reverse-hybridisation assay. 297517 to 297997 represent specifically claimed probes for use in the assay, and 297479 to 297501 represent specifically claimed HIV protease gene polymorphic nucleotide sequences. 297502 to 297515, and 298004 to 298007, represent PCR primers for the HIV protease gene, and 297516 represents an HIV protease probe used in an example from the present invention. The method, probes and primers can be used for the detection of drug-selected mutations in the HIV protease gene. The method allows the simultaneous characterisation of a range of codons involved in drug resistance. The method may also be used for HIV protease genotyping assays. The probes are able to discriminate between wild type and mutated protease sequences. The method allows rapid and reliable detection of drug-selected mutation in HIV.

XX
SQ

Sequence 13 BP; 5 A; 1 C; 2 G; 5 T; 0 other;

Query Match 100.0%; Score 7; DB 21; Length 13;

Best Local Similarity 100.0%; Pred. No. 1.8e+04;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ttgtgata 7

|||||

DB 12 TTGTGATA 6

RESULT 8

V49155

ID V49155 standard; DNA; 14 BP.

XX
AC

V49155;

XX
DT

15-OCT-1998 (first entry)

XX
XX

rb gene antisense oligonucleotide rb-N-103.

XX
DE

rb gene antisense oligonucleotide rb-N-103.

XX

KW rb gene; antisense oligonucleotide; modulate; gene expression; ss.

XX
OS

Synthetic.

XX
OS

Homo sapiens.

XX
PN

EP856579-A1.

XX
XX

05-AUG-1998.

XX
PF

31-JAN-1997; 97EP-0101531.

XX
XX

31-JAN-1997; 97EP-0101531.

XX
PA

(BIOG-) BIOGNOSTIK GES BIOMOLEKULARE DIAGNOSTIK.

XX
PI

Brysch W, Schlingensiepen K;

XX
WPI;

1998-400910/35.

XX
PT

Preparation of antisense oligo:nucleotide(s) which lack long runs of

XX
PT

consecutive guanosine or inosine - and have specific ratio of

XX
PT

residues able to form two or three hydrogen bonds, have greater

XX
PT

activity and reduced toxicity, used therapeutically or to modulate

XX
PT

growth of cells in culture

XX
PS

Example 7; Fig 9c; 286pp; English.

XX
CC

V49008-236 represent antisense oligonucleotides directed against

XX
CC

the rb gene. Of these, only oligonucleotides V49008-52 resulted in

XX
CC

effective downregulation of negative growth control by rb, while

XX
CC

oligonucleotides V49052-236 had little effect. The oligonucleotides

XX
CC

exemplify the invention. The specification describes oligonucleotides

XX
CC

that contain 8-30 nucleotides, which contain at most 8 nucleotides

XX
CC

that can each form three hydrogen bonds to cytosine; do not contain

XX
CC

four consecutive nucleotides able to form three H-bonds each to four

XX
CC

nucleotides each able to form three H-bonds to three consecutive

XX
CC

cytosines, and the ratio between residues able to form two H-bonds

XX
CC

each (2R) or three such bonds (3R) is given by 2R/3R = 0.33-0.72. The

XX
CC

oligonucleotides are used to modulate expression of genes, particularly

XX
CC

the genes for p53, Erb-2, JunB, JunD, TGF-beta 1 or beta 2 to control

XX
CC

proliferation of primary cell cultures (e.g. bone marrow stem, liver or

XX
CC

kidney cells, osteoclasts, osteoblasts and/or keratinocytes). The

XX
CC

oligonucleotides can also be used to analyse function of proteins (by

XX
CC

altering their expression or activity) and therapeutically, e.g. in

XX
CC

cases of cancer or (targeting TGF) for stimulating the immune system.

XX
SQ

Sequence 14 BP; 5 A; 1 C; 1 G; 7 T; 0 other;

XX
SQ

Query Match 100.0%; Score 7; DB 19; Length 14;

Best Local Similarity 100.0%; Pred. No. 1.8e+04;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ttgtgata 7

|||||

DB 1 ttgtgata 7

RESULT 9

297691/c

ID 297691 standard; DNA; 14 BP.

XX
AC

297691;

XX
DT

26-APR-2000 (first entry)

XX
DE

HIV-1 protease gene probe SEQ ID NO:181.

XX
KW

Human immunodeficiency virus; HIV; protease; probe; detection;

XX
KW

drug selected mutation; hybridisation; genotyping; infection;

XX
KW

drug resistance; ss.

OS Human immunodeficiency virus type 1.
 PN WO9967428-A2.
 XX
 XX
 PD 29-DEC-1999.
 XX
 XX
 PF 22-JUN-1999; 99WO-EP04317.
 XX
 XX
 PR 24-JUN-1998; 98EP-0870143.
 XX
 XX
 PA (INNO-) INNOGENETICS NV.
 XX
 PI Stuyver L;
 XX
 XX
 DR WPI; 2000-147219/13.
 XX
 XX
 PT Detection of drug-selected mutations in the HIV protease gene used to
 PT treat HIV infections
 XX
 XX
 PS Claim 3; Page 37; 76pp; English.
 XX
 XX
 CC The present invention describes the detection of drug-selected mutations
 CC in the HIV protease gene. The method of detection allows the simultaneous
 CC characterisation of a range of codons involved in drug resistance using
 CC sets of probes optimised to function together in a reverse-hybridisation
 CC assay. 297517 to 297997 represent specifically claimed probes for use in
 CC the assay, and 297479 to 297501 represent specifically claimed HIV
 CC protease gene polymorphic nucleotide sequences. 297502 to 297515, and
 CC 298004 to 298007, represent PCR primers for the HIV protease gene, and
 CC 297516 represents an HIV protease probe used in an example from the
 CC present invention. The method, probes and primers can be used for the
 CC detection of drug-selected mutations in the HIV protease gene. The
 CC method allows the simultaneous characterisation of a range of codons
 CC involved in drug resistance. The method may also be used for HIV
 CC protease genotyping assays. The probes are able to discriminate between
 CC wild type and mutated protease sequences. The method allows rapid and
 CC reliable detection of drug-selected mutation in HIV.
 XX
 XX
 SQ Sequence 14 BP; 5 A; 1 C; 2 G; 6 T; 0 other;
 Query Match 100.0%; Score 7; DB 21; Length 14;
 Best Local Similarity 100.0%; Pred. No. 1.8e+04;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ttgtgata 7
 Db |||||
 11 TTTGATA 5
 RESULT 10
 Q52243
 ID Q52243 standard; RNA; 15 BP.
 XX
 AC Q52243;
 XX
 XX
 DT 26-MAY-1994 (first entry)
 XX
 XX
 DE Neuroblastoma specific mRNA ribozyme cleavable nucleotide (2670).
 XX
 KW Multiple drug resistance; mdr-1; ribozyme; membrane protein; liver;
 KW resistance; chemotherapeutic agent; colchicine; doxorubicin; colon;
 KW actinomycin D; vinblastine; small intestine; kidney; adrenal gland;
 KW adenocarcinoma; bowel; transformed phenotype; promyelocytic leukemia;
 KW human; chronic myelogenous leukemia; CML; follicular lymphoma;
 KW B-cell acute lymphocytic leukemia; breast cancer; colon carcinoma;
 KW neuroblastoma; lung cancer; genetic drift; mutation; hammerhead motif;
 KW hairpin; hepatitis delta virus; group I intron; RNaseP; leukaemia; ss.
 XX
 OS Homo sapiens.
 XX
 XX
 PN WO93232057-A.
 XX

PD 25-NOV-1993.
 XX
 PF 13-MAY-1993; 93WO-US04573.
 XX
 XX
 PR 14-MAY-1992; 92US-0882822.
 PR 14-MAY-1992; 92US-0882885.
 PR 26-AUG-1992; 92US-0936110.
 PR 26-AUG-1992; 92US-0936421.
 PR 26-AUG-1992; 92US-0936422.
 PR 26-AUG-1992; 92US-0936531.
 PR 26-AUG-1992; 92US-0936532.
 PR 07-DEC-1992; 92US-0987131.
 PR 19-JAN-1993; 93US-0006122.
 PR 19-JAN-1993; 93US-0008910.
 XX
 PA (RIBO-) RIBOZYME PHARM INC.
 XX
 XX
 PI Draper KG, Thompson JD;
 XX
 XX
 DR WPI; 1993-386203/48.
 XX
 XX
 PS Claim 3; Fig 10; 69pp; English.
 XX
 CC The sequences given in Q51825-2266 represent areas of mRNAs which are
 CC associated with development or maintenance of chronic myelogenous
 CC leukemia (CML), promyelocytic leukemia, Burkitt's lymphoma, or
 CC acute lymphocytic leukemia, follicular lymphoma, B-cell acute
 CC lymphocytic leukemia, breast cancer, colon carcinoma, neuroblastoma
 CC and lung cancer. The full length mRNAs containing these target
 CC sequences, encode aberrant cellular proteins which are able to control
 CC cellular proliferation and are directly linked to a leukemic
 CC phenotype. These target sequences are identified by the ribozyme of
 CC the invention. The ribozymes is formed in a hammerhead motif, but may
 CC also be formed in the motif of a hairpin, hepatitis delta virus, group
 CC I intron or RNaseP-like RNA. These ribozymes may be used to inhibit
 CC the development or expression of a transformed phenotype in man and
 CC other animals by modulating expression of the corresponding gene.
 CC Cleavage of target mRNAs expressed in pre-neoplastic and transformed
 CC cells elicits inhibition of the transformed state. Multiple drug
 CC resistance (mdr-1) mRNA specific ribozymes remove the mechanism of
 CC drug resistance used by transformed cells and thus enhances drug
 CC therapies for tumours. The ribozymes may also be used to study
 CC genetic drift and mutations within cells.
 XX
 XX
 SQ Sequence 15 BP; 4 A; 1 C; 1 G; 9 U; 0 other;
 Query Match 100.0%; Score 7; DB 14; Length 15;
 Best Local Similarity 42.9%; Pred. No. 1.8e+04;
 Matches 3; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ttgtgata 7
 Db ::|||:|
 1 uuugaua 7
 RESULT 11
 T56960
 ID T56960 standard; RNA; 15 BP.
 XX
 AC T56960;
 XX
 XX
 DT 24-APR-1997 (first entry)
 XX
 XX
 DE RSV 1C hammerhead ribozyme target sequence (nt. position 16).
 XX
 XX
 KW Enzymatic nucleic acid; ribozyme; trans cleavage; inhibition;
 KW gene expression; downregulation; interleukin-5; IL-5; ICAM-1;
 KW intercellular adhesion molecule; rel A; tumour necrosis factor;

KW TNF-alpha; respiratory syncytial virus; RSV; bcr-abl; oncogene;
 KW translocation; chronic myelogenous leukaemia; CML; cancer;
 KW Philadelphia chromosome; inflammation; autoimmune disease;
 KW atherosclerosis; myocardial infarction; stroke; restenosis;
 KW transplant rejection; rheumatoid arthritis; psoriasis;
 KW myocardial ischaemia; Kawasaki disease; septic shock; HIV;
 KW human immunodeficiency virus; acquired immune deficiency syndrome;
 KW AIDS; ss.

XX Respiratory Syncytial Virus.

XX WO9523225-A2.

XX 31-AUG-1995.

XX 23-FEB-1995; 95WO-IB00156.

XX 30-JAN-1995; 95US-0380734.

XX 23-FEB-1994; 94US-0201109.

XX 29-MAR-1994; 94US-0218934.

XX 04-APR-1994; 94US-0222795.

XX 07-APR-1994; 94US-0224483.

XX 15-APR-1994; 94US-0227958.

XX 15-APR-1994; 94US-0228041.

XX 18-MAY-1994; 94US-0245736.

XX 06-JUL-1994; 94US-0271280.

XX 15-AUG-1994; 94US-0291932.

XX 16-AUG-1994; 94US-0291433.

XX 17-AUG-1994; 94US-0292620.

XX 19-AUG-1994; 94US-0293520.

XX 02-SEP-1994; 94US-0300000.

XX 08-SEP-1994; 94US-0303039.

XX 23-SEP-1994; 94US-0311486.

XX 23-SEP-1994; 94US-0311749.

XX 28-SEP-1994; 94US-0314397.

XX 03-OCT-1994; 94US-0316771.

XX 07-OCT-1994; 94US-0319492.

XX 11-OCT-1994; 94US-0321993.

XX 04-NOV-1994; 94US-0334847.

XX 10-NOV-1994; 94US-0337608.

XX 28-NOV-1994; 94US-0345516.

XX 16-DEC-1994; 94US-0357577.

XX 23-DEC-1994; 94US-0363233.

XX (RIBO-) RIBOZYME PHARM INC.

XX Stinchcomb DT, Chowrira B, Drenzo A, Draper KG, Dudycz LW;

PI Grimm S, Karpeisky A, Kisich K, Matulic-Adamic J;

PI McSwiggen JA, Modak A, Pavco P, Beigelman L, Sullivan SM;

PI Sweedler D, Thompson JD, Tracz D, Usman N, Wincott FE;

PI Woolf T;

XX WPI; 1995-351090/45.

XX Ribozymes having modified bases and methods for producing them

XX for use in inhibiting disease related genes

XX Claim 2: Page 269; 407pp: English.

XX The present sequence represents a preferred target sequence for an

XX enzymatic nucleic acid (i.e. a ribozyme) which cleaves mRNA coding

XX for a protein of respiratory syncytial virus (RSV) at the

XX nucleotide base position indicated in the DE line. Regions of

XX the mRNA that do not form secondary folding structures and that

XX contain potential hammerhead and hairpin ribozyme cleavage sites

XX were identified by computer analysis. Ribozymes directed against

XX these mRNA sequences were designed and synthesised with modifications

XX that improve their nuclease resistance. The ribozymes cleave the

XX target sequences and can be used for treatment and diagnosis of

XX RSV infection.

XX Sequence 15 BP; 7 A; 0 C; 3 G; 5 U; 0 other;

SQ

Query Match 100.0%; Score 7; DB 16; Length 15;
 Best Local Similarity 42.9%; Pred. No. 1.8e+04;
 Matches 3; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtgata 7

Db 7 uuugaua 13

RESULT 12

T56962

ID T56962 standard; RNA; 15 BP.

XX

AC T56962;

XX

DT 24-APR-1997 (first entry)

XX

DE RSV 1C hammerhead ribozyme target sequence (nt. position 17).

XX

KW Enzymatic nucleic acid; ribozyme; trans cleavage; inhibition;
 KW gene expression; downregulation; interleukin-5; IL-5; ICAM-1;
 KW intercellular adhesion molecule; rel A; tumour necrosis factor;
 KW TNF-alpha; respiratory syncytial virus; RSV; bcr-abl; oncogene;
 KW translocation; chronic myelogenous leukaemia; CML; cancer;
 KW Philadelphia chromosome; inflammation; autoimmune disease;
 KW atherosclerosis; myocardial infarction; stroke; restenosis;
 KW transplant rejection; rheumatoid arthritis; psoriasis;
 KW myocardial ischaemia; Kawasaki disease; septic shock; HIV;
 KW human immunodeficiency virus; acquired immune deficiency syndrome;
 KW AIDS; ss.

XX

OS Respiratory Syncytial Virus.

XX

PN WO9523225-A2.

XX

PD 31-AUG-1995.

XX

PF 23-FEB-1995; 95WO-IB00156.

XX

PR 30-JAN-1995; 95US-0380734.

PR

PR 23-FEB-1994; 94US-0201109.

PR

PR 29-MAR-1994; 94US-0218934.

PR

PR 04-APR-1994; 94US-0222795.

PR

PR 07-APR-1994; 94US-0224483.

PR

PR 15-APR-1994; 94US-0227958.

PR

PR 15-APR-1994; 94US-0228041.

PR

PR 18-MAY-1994; 94US-0245736.

PR

PR 06-JUL-1994; 94US-0271280.

PR

PR 15-AUG-1994; 94US-0291932.

PR

PR 16-AUG-1994; 94US-0291433.

PR

PR 17-AUG-1994; 94US-0292620.

PR

PR 19-AUG-1994; 94US-0293520.

PR

PR 02-SEP-1994; 94US-0300000.

PR

PR 08-SEP-1994; 94US-0303039.

PR

PR 23-SEP-1994; 94US-0311486.

PR

PR 23-SEP-1994; 94US-0311749.

PR

PR 28-SEP-1994; 94US-0314397.

PR

PR 03-OCT-1994; 94US-0316771.

PR

PR 07-OCT-1994; 94US-0319492.

PR

PR 11-OCT-1994; 94US-0321993.

PR

PR 04-NOV-1994; 94US-0334847.

PR

PR 10-NOV-1994; 94US-0337608.

PR

PR 28-NOV-1994; 94US-0345516.

PR

PR 16-DEC-1994; 94US-0357577.

PR

PR 23-DEC-1994; 94US-0363233.

PR

XX (RIBO-) RIBOZYME PHARM INC.

XX Stinchcomb DT, Chowrira B, Drenzo A, Draper KG, Dudycz LW;

PI Grimm S, Karpeisky A, Kisich K, Matulic-Adamic J;

PI McSwiggen JA, Modak A, Pavco P, Beigelman L, Sullivan SM;

PI Sweedler D, Thompson JD, Tracz D, Usman N, Wincott FE;

```
PI Woolf T;
XX
DR WPI; 1995-351090/45.
XX
XX Ribozymes having modified bases and methods for producing them
PT for use in inhibiting disease related genes
PT
XX
XX Claim 2; Page 269; 407pp; English.
XX
XX The present sequence represents a preferred target sequence for an
CC enzymatic nucleic acid (i.e. a ribozyme) which cleaves mRNA coding
CC for a protein of respiratory syncytial virus (RSV) at the
CC nucleotide base position indicated in the DE line. Regions of
CC the mRNA that do not form secondary folding structures and that
CC contain potential hammerhead and hairpin ribozyme cleavage sites
CC were identified by computer analysis. Ribozymes directed against
CC these mRNA sequences were designed and synthesised with modifications
CC that improve their nuclease resistance. The ribozymes cleave the
CC target sequences and can be used for treatment and diagnosis of
CC RSV infection.
XX
XX Sequence 15 BP; 7 A; 0 C; 3 G; 5 U; 0 other;
SQ
Query Match 100.0%; Score 7; DB 16; Length 15;
Best Local Similarity 42.9%; Pred. No. 1.8e+04;
Matches 3; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
QY 1 ttgtgata 7
DB 6 uuugaua 12
RESULT 13
ID T56964 standard; RNA; 15 BP.
XX
XX T56964;
XX
XX 24-APR-1997 (first entry)
DT
XX
XX RSV 1C hammerhead ribozyme target sequence (nt. position 21).
DE
XX
XX Enzymatic nucleic acid; ribozyme; trans cleavage; inhibition;
KW gene expression; downregulation; interleukin-5; IL-5; ICAM-1;
KW intercellular adhesion molecule; rel A; tumour necrosis factor;
KW TNF-alpha; respiratory syncytial virus; RSV; bcr-abl; oncogene;
KW translocation; chronic myelogenous leukaemia; CML; cancer;
KW Philadelphia chromosome; inflammation; autoimmune disease;
KW atherosclerosis; myocardial infarction; stroke; restenosis;
KW transplant rejection; rheumatoid arthritis; psoriasis;
KW myocardial ischaemia; Kawasaki disease; septic shock; HIV;
KW human immunodeficiency virus; acquired immune deficiency syndrome;
KW AIDS; ss.
XX
XX Respiratory Syncytial Virus.
OS
XX
XX W09523225-A2.
PN
XX
XX 31-AUG-1995.
PD
XX
XX 23-FEB-1995; 95WO-IB00156.
PF
XX
XX 30-JAN-1995; 95US-0380734.
PR
XX 23-FEB-1994; 94US-0201109.
PR
XX 29-MAR-1994; 94US-0218934.
PR
XX 04-APR-1994; 94US-0222795.
PR
XX 07-APR-1994; 94US-0224483.
PR
XX 15-APR-1994; 94US-0227958.
PR
XX 18-MAY-1994; 94US-0228041.
PR
XX 06-JUL-1994; 94US-0245736.
PR
XX 15-AUG-1994; 94US-0271280.
PR
XX 15-AUG-1994; 94US-0291932.
PR
XX 16-AUG-1994; 94US-0291433.
PR
XX 17-AUG-1994; 94US-0292620.
PR
XX 19-AUG-1994; 94US-0293520.
PR
XX 02-SEP-1994; 94US-0300000.
PR
XX 08-SEP-1994; 94US-0303039.
PR
XX 23-SEP-1994; 94US-0311486.
PR
XX 28-SEP-1994; 94US-0311749.
PR
XX 03-OCT-1994; 94US-0314397.
PR
XX 07-OCT-1994; 94US-0316771.
PR
XX 11-OCT-1994; 94US-0319492.
PR
XX 04-NOV-1994; 94US-0321993.
PR
XX 28-NOV-1994; 94US-0334847.
PR
XX 16-DEC-1994; 94US-0337608.
PR
XX 23-DEC-1994; 94US-0345516.
PR
XX (RIBO-) RIBOZYME PHARM INC.
PA
XX Stinchcomb DT, Chowrira B, Drenzo A, Draper KG, Dudycz LM;
XX Grimm S, Karpelsky A, Kisich K, Matulic-Adamic J;
XX McSwiggen JA, Modak A, Pavco P, Beigelman L, Sullivan SM;
XX Sweedler D, Thompson JD, Tracz D, Usman N, Wincott FE;
XX Woolf T;
XX
XX WPI; 1995-351090/45.
XX
XX Ribozymes having modified bases and methods for producing them
PT for use in inhibiting disease related genes
PT
XX
XX Claim 2; Page 269; 407pp; English.
XX
XX The present sequence represents a preferred target sequence for an
CC enzymatic nucleic acid (i.e. a ribozyme) which cleaves mRNA coding
CC for a protein of respiratory syncytial virus (RSV) at the
CC nucleotide base position indicated in the DE line. Regions of
CC the mRNA that do not form secondary folding structures and that
CC contain potential hammerhead and hairpin ribozyme cleavage sites
CC were identified by computer analysis. Ribozymes directed against
CC these mRNA sequences were designed and synthesised with modifications
CC that improve their nuclease resistance. The ribozymes cleave the
CC target sequences and can be used for treatment and diagnosis of
CC RSV infection.
XX
XX Sequence 15 BP; 6 A; 2 C; 2 G; 5 U; 0 other;
SQ
Query Match 100.0%; Score 7; DB 16; Length 15;
Best Local Similarity 42.9%; Pred. No. 1.8e+04;
Matches 3; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
QY 1 ttgtgata 7
DB 2 uuugaua 8
RESULT 14
ID T57424 standard; RNA; 15 BP.
XX
XX T57424;
XX
XX 19-MAR-1997 (first entry)
DT
XX
XX RSV N hammerhead ribozyme target sequence (nt. position 1146).
DE
XX
XX Enzymatic nucleic acid; ribozyme; trans cleavage; inhibition;
KW gene expression; downregulation; interleukin-5; IL-5; ICAM-1;
KW intercellular adhesion molecule; rel A; tumour necrosis factor;
KW TNF-alpha; respiratory syncytial virus; RSV; bcr-abl; oncogene;
KW translocation; chronic myelogenous leukaemia; CML; cancer;
KW Philadelphia chromosome; inflammation; autoimmune disease;
KW atherosclerosis; myocardial infarction; stroke; restenosis;
```

KW transplant rejection; rheumatoid arthritis; psoriasis;
 KW myocardial ischaemia; Kawasaki disease; septic shock; HIV;
 KW human immunodeficiency virus; acquired immune deficiency syndrome;
 KW AIDS; ss.

XX Respiratory Syncytial Virus.

OS WO9523225-A2.

XX 31-AUG-1995.

XX 23-FEB-1995; 95WO-IB00156.

XX 30-JAN-1995; 95US-0380734.

PR 23-FEB-1994; 94US-0201109.

PR 29-MAR-1994; 94US-0218934.

PR 04-APR-1994; 94US-0222795.

PR 07-APR-1994; 94US-0224483.

PR 15-APR-1994; 94US-0227958.

PR 15-APR-1994; 94US-0228041.

PR 18-MAY-1994; 94US-0245736.

PR 06-JUL-1994; 94US-0271280.

PR 15-AUG-1994; 94US-0291932.

PR 16-AUG-1994; 94US-0291433.

PR 17-AUG-1994; 94US-0292620.

PR 19-AUG-1994; 94US-0293520.

PR 02-SEP-1994; 94US-0300000.

PR 08-SEP-1994; 94US-0303039.

PR 23-SEP-1994; 94US-0311486.

PR 23-SEP-1994; 94US-0311749.

PR 28-SEP-1994; 94US-0314397.

PR 03-OCT-1994; 94US-0316771.

PR 07-OCT-1994; 94US-0319492.

PR 11-OCT-1994; 94US-0321993.

PR 04-NOV-1994; 94US-0334847.

PR 10-NOV-1994; 94US-0337608.

PR 28-NOV-1994; 94US-0345516.

PR 16-DEC-1994; 94US-0357577.

PR 23-DEC-1994; 94US-0363233.

XX (RIBO-) RIBOZYME PHARM INC.

XX Stinchcomb DT, Chowira B, Drenzo A, Draper KG, Dudycz LW;

PI Grimm S, Karpeisky A, Kisich K, Matulic-Adamic J;

PI McSwiggen JA, Modak A, Pavco P, Belgelman L, Sullivan SM;

PI Sweedler D, Thompson JD, Tracz D, Usman N, Wincott FE;

PI Woolf T;

XX WPI; 1995-351090/45.

XX RIBOZYME PHARM INC.

XX Stinchcomb DT, Chowira B, Drenzo A, Draper KG, Dudycz LW;
 PI Grimm S, Karpeisky A, Kisich K, Matulic-Adamic J;
 PI McSwiggen JA, Modak A, Pavco P, Belgelman L, Sullivan SM;
 PI Sweedler D, Thompson JD, Tracz D, Usman N, Wincott FE;
 PI Woolf T;
 XX WPI; 1995-351090/45.
 XX Ribozymes having modified bases and methods for producing them -
 PT for use in inhibiting disease related genes.
 XX Claim 2; Page 276; 407pp; English.
 XX The present sequence represents a preferred target sequence for an
 CC enzymatic nucleic acid (i.e. a ribozyme) which cleaves mRNA coding
 CC for a protein of respiratory syncytial virus (RSV) at the
 CC nucleotide base position indicated in the DE line. Regions of
 CC the mRNA that do not form secondary folding structures and that
 CC contain potential hammerhead and hairpin ribozyme cleavage sites
 CC were identified by computer analysis. Ribozymes directed against
 CC these mRNA sequences were designed and synthesised with modifications
 CC that improve their nuclease resistance. The ribozymes cleave the
 CC target sequences and can be used for treatment and diagnosis of
 CC RSV infection.
 XX Sequence 15 BP; 7 A; 3 C; 2 U; 0 other;

Query Match 100.0%; Score 7; DB 16; Length 15;

Best Local Similarity 100.0%; Pred. No. 1.8e+04;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ttgtgata 7
 Db 13 TTTGATA 7
 RESULT 15
 T57425/c
 ID T57425 standard; RNA; 15 BP.
 XX
 AC T57425;
 XX
 DT 19-MAR-1997 (first entry)
 XX
 DE RSV N hammerhead ribozyme target sequence (nt. position 1148).
 XX
 KW Enzymatic nucleic acid; ribozyme; trans cleavage; inhibition;
 KW gene expression; downregulation; interleukin-5; IL-5; ICAM-1;
 KW intercellular adhesion molecule; rel A; tumour necrosis factor;
 KW TNF-alpha; respiratory syncytial virus; RSV; bcr-abl; oncogene;
 KW translocation; chronic myelogenous leukaemia; CML; cancer;
 KW Philadelphia chromosome; inflammation; autoimmune disease;
 KW atherosclerosis; myocardial infarction; stroke; restenosis;
 KW transplant rejection; rheumatoid arthritis; psoriasis;
 KW myocardial ischaemia; Kawasaki disease; septic shock; HIV;
 KW human immunodeficiency virus; acquired immune deficiency syndrome;
 KW AIDS; ss.
 XX
 OS Respiratory Syncytial Virus.
 XX
 PN WO9523225-A2.
 XX
 PD 31-AUG-1995.
 XX
 PF 23-FEB-1995; 95WO-IB00156.
 XX
 PR 30-JAN-1995; 95US-0380734.
 PR 23-FEB-1994; 94US-0201109.
 PR 29-MAR-1994; 94US-0218934.
 PR 04-APR-1994; 94US-0222795.
 PR 07-APR-1994; 94US-0224483.
 PR 15-APR-1994; 94US-0227958.
 PR 15-APR-1994; 94US-0228041.
 PR 18-MAY-1994; 94US-0245736.
 PR 06-JUL-1994; 94US-0271280.
 PR 15-AUG-1994; 94US-0291932.
 PR 16-AUG-1994; 94US-0291433.
 PR 17-AUG-1994; 94US-0292620.
 PR 19-AUG-1994; 94US-0293520.
 PR 02-SEP-1994; 94US-0300000.
 PR 08-SEP-1994; 94US-0303039.
 PR 23-SEP-1994; 94US-0311486.
 PR 23-SEP-1994; 94US-0311749.
 PR 28-SEP-1994; 94US-0314397.
 PR 03-OCT-1994; 94US-0316771.
 PR 07-OCT-1994; 94US-0319492.
 PR 11-OCT-1994; 94US-0321993.
 PR 04-NOV-1994; 94US-0334847.
 PR 10-NOV-1994; 94US-0337608.
 PR 28-NOV-1994; 94US-0345516.
 PR 16-DEC-1994; 94US-0357577.
 PR 23-DEC-1994; 94US-0363233.
 XX
 PA (RIBO-) RIBOZYME PHARM INC.
 XX
 PI Stinchcomb DT, Chowira B, Drenzo A, Draper KG, Dudycz LW;
 PI Grimm S, Karpeisky A, Kisich K, Matulic-Adamic J;
 PI McSwiggen JA, Modak A, Pavco P, Belgelman L, Sullivan SM;
 PI Sweedler D, Thompson JD, Tracz D, Usman N, Wincott FE;
 PI Woolf T;
 XX WPI; 1995-351090/45.

PT Ribozymes having modified bases and methods for producing them
PT , for use in inhibiting disease related genes
XX
XX Claim 2; Page 276; 407pp; English.
XX
CC The present sequence represents a preferred target sequence for an
CC enzymatic nucleic acid (i.e. a ribozyme) which cleaves mRNA coding
CC for a protein of respiratory syncytial virus (RSV) at the
CC nucleotide base position indicated in the DE line. Regions of
CC the mRNA that do not form secondary folding structures and that
CC contain potential hammerhead and hairpin ribozyme cleavage sites
CC were identified by computer analysis. Ribozymes directed against
CC these mRNA sequences were designed and synthesised with modifications
CC that improve their nuclease resistance. The ribozymes cleave the
CC target sequences and can be used for treatment and diagnosis of
CC RSV infection.
XX
SQ Sequence 15 BP; 6 A; 4 C; 2 G; 3 U; 0 other;

Query Match 100.0%; Score 7; DB 16; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.8e+04;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ttgtata 7
| | | | |
Db 11 TTTGATA 5

Search completed: March 27, 2001, 08:22:59
Job time: 5301 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 27, 2001, 08:19:22 ; Search time 132.3 Seconds
(without alignments)
8.527 Million cell updates/sec

Title: US-09-380-826A-5

Perfect score: 7

Sequence: 1 ttgtgata 7

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 280836 seqs, 80580151 residues

Total number of hits satisfying chosen parameters: 561672

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents.NA.*

1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*

2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*

3: /cgn2_6/ptodata/2/ina/6_COMB.seq.*

4: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*

5: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	7	100.0	11	1	US-08-086-634-6
C 2	7	100.0	11	1	US-08-086-634-7
C 3	7	100.0	13	1	US-08-166-664-5
C 4	7	100.0	13	1	US-08-271-880A-217
C 5	7	100.0	13	2	US-08-910-408-217
C 6	7	100.0	13	3	US-09-249-213-217
C 7	7	100.0	14	1	US-08-271-880A-212
C 8	7	100.0	14	2	US-08-910-408-212
C 9	7	100.0	14	3	US-09-249-215-212
C 10	7	100.0	15	1	US-08-373-124A-90
C 11	7	100.0	15	1	US-08-291-932A-129
C 12	7	100.0	15	1	US-08-291-932A-130
C 13	7	100.0	15	1	US-08-291-932A-131
C 14	7	100.0	15	1	US-08-334-847-8
C 15	7	100.0	15	1	US-08-334-847-9
C 16	7	100.0	15	1	US-08-334-847-10
C 17	7	100.0	15	1	US-08-334-847-655
C 18	7	100.0	15	1	US-08-334-847-656
C 19	7	100.0	15	1	US-08-435-628-90
C 20	7	100.0	15	2	US-08-585-684B-113
C 21	7	100.0	15	2	US-08-585-684B-114
C 22	7	100.0	15	2	US-08-585-684B-115
C 23	7	100.0	15	2	US-08-585-684B-116
C 24	7	100.0	15	2	US-08-585-684B-2126
C 25	7	100.0	15	2	US-08-585-684B-2127
C 26	7	100.0	15	2	US-08-585-684B-2128
C 27	7	100.0	15	2	US-08-585-684B-2129
C 28	7	100.0	15	2	US-08-585-684B-2319

ALIGNMENTS

RESULT 1

US-08-086-634-6/c

; Sequence 6, Application US/08086634

; Patent No. 5516634

; GENERAL INFORMATION:

; APPLICANT: NEWMAN, Peter J.

; APPLICANT: SANTOSO, Sentot Saleh

; TITLE OF INVENTION: Molecular Basis of the

; TITLE OF INVENTION: Human Platelet Bra/Brb Alloantigen System and

; NUMBER OF SEQUENCES: 12

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Foley & Lardner

; STREET: P. O. Box 1497

; CITY: Madison

; STATE: Wisconsin

; COUNTRY: USA

; ZIP: 53701-1497

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy Disk

; COMPUTER: IBM PC Compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: WordPerfect, Version 5.1

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/086,634

; FILING DATE: 30-June-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Scanlon, William J.

; REGISTRATION NUMBER: 30,136

; REFERENCE/DOCKET NUMBER: 30383/132

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (608) 258-4284

; TELEFAX: (608) 258-4258

; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 11 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA to mRNA

US-08-086-634-6

Query Match 100.0%; Score 7; DB 1; Length 11;

Best Local Similarity 100.0%; Pred No. 7.1e+03;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtgata 7

Db 7 TTTGATA 1

Sequence 2320, Ap
Sequence 2321, Ap
Sequence 25, Appl
Sequence 7, Appl
Sequence 8, Appl
Sequence 520, App
Sequence 522, App
Sequence 1561, Ap
Sequence 1563, Ap
Sequence 2411, Ap
Sequence 2413, Ap
Sequence 2565, Ap
Sequence 2567, Ap
Sequence 520, App
Sequence 522, App
Sequence 1561, Ap
Sequence 1563, Ap

RESULT 2
US-08-086-634-7/c
; Sequence 7, Application US/08086634
; Patent No. 5516634
; GENERAL INFORMATION:
; APPLICANT: NEWMAN, Peter J.
; APPLICANT: SANTOSO, Sentot Saleh
; TITLE OF INVENTION: Molecular Basis of the
; TITLE OF INVENTION: Human Platelet Bra/Brb Alloantigen System and
; TITLE OF INVENTION: Applications Thereof
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: P. O. Box 1497
; CITY: Madison
; STATE: Wisconsin
; COUNTRY: USA
; ZIP: 53701-1497
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect, Version 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/086.634
; FILING DATE: 30-June-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Scanlon, William J.
; REGISTRATION NUMBER: 30136
; REFERENCE/DOCKET NUMBER: 30383/132
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 258-4284
; TELEFAX: (608) 258-4258
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
US-08-086-634-7

Query Match 100.0%; Score 7; DB 1; Length 11;
Best Local Similarity 100.0%; Pred. No. 7.1e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtgata 7
| | | | |
DB 7 TTTGATA 1

RESULT 3
US-08-166-664-5/c
; Sequence 5, Application US/08166664
; Patent No. 5646020
; GENERAL INFORMATION:
; APPLICANT: James A. McSwiggen
; APPLICANT: J. Anthony Mamone
; TITLE OF INVENTION: HAMMERHEAD RIBOZYMES FOR
; TITLE OF INVENTION: PREFERRED TARGETS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 611 West Sixth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
; COMPUTER: IBM Compatible

; OPERATING SYSTEM: IBM P.C. DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/166,664
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/884,074
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 197/062
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-166-664-5

Query Match 100.0%; Score 7; DB 1; Length 13;
Best Local Similarity 100.0%; Pred. No. 7e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtgata 7
| | | | |
DB 10 TTTGATA 4

RESULT 4
US-08-271-880A-217/c
; Sequence 217, Application US/08271880A
; Patent No. 5693535
; GENERAL INFORMATION:
; APPLICANT: Kenneth G. Draper
; APPLICANT: Bharat Chowhira
; APPLICANT: James McSwiggen
; APPLICANT: Dan T. Stinchcomb
; APPLICANT: James D. Thompson
; TITLE OF INVENTION: METHOD AND REAGENT FOR INHIBITING
; TITLE OF INVENTION: HUMAN IMMUNODEFICIENCY VIRUS
; TITLE OF INVENTION: REPLICATION
; NUMBER OF SEQUENCES: 232
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/271,880A
; FILING DATE: July 7, 1994
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below:
; APPLICATION NUMBER: 08/103,243
; FILING DATE: August 6, 1993
; APPLICATION NUMBER: 07/882,886
; FILING DATE: May 14, 1992

two

ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 206/116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 217:
SEQUENCE CHARACTERISTICS:
LENGTH: 13 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-271-880A-217

Query Match 100.0%; Score 7; DB 1; Length 13;
Best Local Similarity 100.0%; Pred. No. 7e+03; 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtgata 7
DB 10 TTTGATA 4

RESULT 5
US-08-910-408-217/c
Sequence 217, Application US/08910408
Patent No. 5972704
GENERAL INFORMATION:
APPLICANT: Kenneth G. Draper
APPLICANT: Bharat Chowrira
APPLICANT: James McSwiggen
APPLICANT: Dan T. Stinchcomb
APPLICANT: James D. Thompson
TITLE OF INVENTION: METHOD AND REAGENT FOR INHIBITING
TITLE OF INVENTION: HUMAN IMMUNODEFICIENCY VIRUS
NUMBER OF SEQUENCES: 232
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: FastSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/910,408
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/271,880
FILING DATE: July 7, 1994
APPLICATION NUMBER: 08/103,243
FILING DATE: August 6, 1993
APPLICATION NUMBER: 07/882,886
FILING DATE: May 14, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 206/116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 217:

SEQUENCE CHARACTERISTICS:
LENGTH: 13 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-910-408-217

Query Match 100.0%; Score 7; DB 2; Length 13;
Best Local Similarity 100.0%; Pred. No. 7e+03; 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtgata 7
DB 10 TTTGATA 4

RESULT 6
US-09-249-215-217/c
Sequence 217, Application US/09249215
Patent No. 6159692
GENERAL INFORMATION:
APPLICANT: Kenneth G. Draper
APPLICANT: Bharat Chowrira
APPLICANT: James McSwiggen
APPLICANT: Dan T. Stinchcomb
APPLICANT: James D. Thompson
TITLE OF INVENTION: METHOD AND REAGENT FOR INHIBITING
TITLE OF INVENTION: HUMAN IMMUNODEFICIENCY VIRUS
REPLICATION
NUMBER OF SEQUENCES: 232
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: FastSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/249,215
FILING DATE: 12-Feb-1999
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/910,408
FILING DATE: <Unknown>
APPLICATION NUMBER: 08/103,243
FILING DATE: August 6, 1993
APPLICATION NUMBER: 07/882,886
FILING DATE: May 14, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 206/116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 217:
SEQUENCE CHARACTERISTICS:
LENGTH: 13 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 217:
US-09-249-215-217

Wed Mar 28 14:03:22 2001

Query Match 100.0%; Score 7; DB 3; Length 13;
Best Local Similarity 100.0%; Pred. No. 7e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtata 7
DB 10 TTTGATA 4

RESULT 7
US-08-271-880A-212
; Sequence 212, Application US/08271880A
; Patent No. 5693535
; GENERAL INFORMATION:
; APPLICANT: Kenneth G. Draper
; APPLICANT: Bharat Chowrira
; APPLICANT: James McSwiggen
; APPLICANT: Dan T. Stinchcomb
; APPLICANT: James D. Thompson
; TITLE OF INVENTION: METHOD AND REAGENT FOR INHIBITING
; TITLE OF INVENTION: HUMAN IMMUNODEFICIENCY VIRUS
; NUMBER OF SEQUENCES: 232
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Suite 4700
; STATE: Los Angeles
; CITY: California
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: FastSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/271,880A
; FILING DATE: July 7, 1994
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below:
; APPLICATION NUMBER: 08/103,243
; FILING DATE: August 6, 1993
; APPLICATION NUMBER: 07/882,886
; FILING DATE: May 14, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 206/116
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 212:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-271-880A-212

Query Match 100.0%; Score 7; DB 1; Length 14;
Best Local Similarity 100.0%; Pred. No. 7e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtata 7
DB 3 TTTGATA 9

RESULT 8
US-08-910-408-212
; Sequence 212, Application US/08910408
; Patent No. 5972704
; GENERAL INFORMATION:
; APPLICANT: Kenneth G. Draper
; APPLICANT: Bharat Chowrira
; APPLICANT: James McSwiggen
; APPLICANT: Dan T. Stinchcomb
; APPLICANT: James D. Thompson
; TITLE OF INVENTION: METHOD AND REAGENT FOR INHIBITING
; TITLE OF INVENTION: HUMAN IMMUNODEFICIENCY VIRUS
; NUMBER OF SEQUENCES: 232
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Suite 4700
; STATE: Los Angeles
; CITY: California
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: FastSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/910,408
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/271,880
; FILING DATE: July 7, 1994
; APPLICATION NUMBER: 08/103,243
; FILING DATE: August 6, 1993
; APPLICATION NUMBER: 07/882,886
; FILING DATE: May 14, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 206/116
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 212:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-910-408-212

Query Match 100.0%; Score 7; DB 2; Length 14;
Best Local Similarity 100.0%; Pred. No. 7e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtata 7
DB 3 TTTGATA 9

RESULT 9
US-09-249-215-212
; Sequence 212, Application US/09249215
; Patent No. 6159692
; GENERAL INFORMATION:
; APPLICANT: Kenneth G. Draper
; APPLICANT: Bharat Chowrira
; APPLICANT: James McSwiggen
; APPLICANT: Dan T. Stinchcomb

```

; James D. Thompson
; TITLE OF INVENTION: METHOD AND REAGENT FOR INHIBITING
; HUMAN IMMUNODEFICIENCY VIRUS
; REPLICATION
;
; NUMBER OF SEQUENCES: 232
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
;
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: FastSEQ Version 1.5
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/249,215
; FILING DATE: 12-Feb-1999
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/910,408
; FILING DATE: <Unknown>
; APPLICATION NUMBER: 08/103,243
; FILING DATE: August 6, 1993
; APPLICATION NUMBER: 07/882,886
; FILING DATE: May 14, 1992
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 206/116
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
;
; INFORMATION FOR SEQ ID NO: 212:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 212:
;
; US-09-249-215-212

```

```

Query Match 100.0%; Score 7; DB 3; Length 14;
Best Local Similarity 100.0%; Pred. No. 7e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 ttgtgata 7
Db 3 TTTGATA 9

```

```

RESULT 10
US-08-373-124A-90/c
; Sequence 90, Application US/08373124A
; Patent No. 5646042
; GENERAL INFORMATION:
; APPLICANT: Stinchcomb, Dan T.
; APPLICANT: Draper, Kenneth
; APPLICANT: McSwiggen, James
; APPLICANT: Jarvis, Thale
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
; TREATMENT OF RESTENOSIS AND
; TITLE OF INVENTION: CANCER USING RIBOZYMES
; NUMBER OF SEQUENCES: 2627
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; Suite 4700

```

```

; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
;
; MEDIUM TYPE: Storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/373,124A
; FILING DATE: January 13, 1995
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/245,466
; FILING DATE: May 18, 1994
; APPLICATION NUMBER: 08/192,943
; FILING DATE: February 7, 1994
; APPLICATION NUMBER: 07/987,132
; FILING DATE: December 7, 1992
; APPLICATION NUMBER: 07/936,422
; FILING DATE: August 26, 1992
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 209/035
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
;
; INFORMATION FOR SEQ ID NO: 90:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; US-08-373-124A-90

```

```

Query Match 100.0%; Score 7; DB 1; Length 15;
Best Local Similarity 100.0%; Pred. No. 7e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 ttgtgata 7
Db 11 TTTGATA 5

```

```

RESULT 11
US-08-291-932A-129
; Sequence 129, Application US/08291932A
; Patent No. 5658780
; GENERAL INFORMATION:
; APPLICANT: Stinchcomb, Dan T.
; APPLICANT: Draper, Kenneth G.
; APPLICANT: McSwiggen, James
; TITLE OF INVENTION: RIBOZYME TREATMENT OF
; DISEASES OR CONDITIONS
; TITLE OF INVENTION: RELATED TO LEVELS OF
; TITLE OF INVENTION: NF-KB
; NUMBER OF SEQUENCES: 830
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: Storage
; COMPUTER: IBM Compatible

```

```

; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/291.932A
; FILING DATE: August 15, 1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below:
; APPLICATION NUMBER: 08/245,466
; FILING DATE: May 18, 1994
; APPLICATION NUMBER: 07/987,132
; FILING DATE: December 7, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 208/157
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 130:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-291-932A-129

Query Match 100.0%; Score 7; DB 1; Length 15;
Best Local Similarity 42.9%; Pred. No. 7e+03;
Matches 3; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ttgtgata 7
Db 9 UUUGAUA 15

RESULT 12
US-08-291-932A-130
; Sequence 130, Application US/08291932A
; Patent No. 5658780
; GENERAL INFORMATION:
; APPLICANT: Stinchcomb, Dan T.
; APPLICANT: Draper, Kenneth G.
; APPLICANT: McSwiggen, James
; TITLE OF INVENTION: RIBOZYME TREATMENT OF
; TITLE OF INVENTION: DISEASES OR CONDITIONS
; TITLE OF INVENTION: RELATED TO LEVELS OF
; TITLE OF INVENTION: NF-KB
; NUMBER OF SEQUENCES: 830
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/291.932A
; FILING DATE: August 15, 1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below:
; APPLICATION NUMBER: 08/245,466
; FILING DATE: May 18, 1994
; APPLICATION NUMBER: 07/987,132
; FILING DATE: December 7, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 208/157
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 129:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-291-932A-130

```

```

; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/291.932A
; FILING DATE: August 15, 1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below:
; APPLICATION NUMBER: 08/245,466
; FILING DATE: May 18, 1994
; APPLICATION NUMBER: 07/987,132
; FILING DATE: December 7, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 208/157
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 129:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-291-932A-129

Query Match 100.0%; Score 7; DB 1; Length 15;
Best Local Similarity 42.9%; Pred. NO. 7e+03;
Matches 3; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ttgtgata 7
Db 9 UUUGAUA 15

RESULT 12
US-08-291-932A-130
; Sequence 130, Application US/08291932A
; Patent No. 5658780
; GENERAL INFORMATION:
; APPLICANT: Stinchcomb, Dan T.
; APPLICANT: Draper, Kenneth G.
; APPLICANT: McSwiggen, James
; TITLE OF INVENTION: RIBOZYME TREATMENT OF
; TITLE OF INVENTION: DISEASES OR CONDITIONS
; TITLE OF INVENTION: RELATED TO LEVELS OF
; TITLE OF INVENTION: NF-KB
; NUMBER OF SEQUENCES: 830
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/291.932A
; FILING DATE: August 15, 1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below:
; APPLICATION NUMBER: 08/245,466
; FILING DATE: May 18, 1994
; APPLICATION NUMBER: 07/987,132
; FILING DATE: December 7, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 208/157
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 130:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-291-932A-130

```

TELEPHONE: (213) 489-1600
 TELEFAX: (213) 955-0440
 TELEX: 67-3510
 INFORMATION FOR SEQ ID NO: 131:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 15 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-291-932A-131

Query Match 100.0%; Score 7; DB 1; Length 15;
 Best Local Similarity 42.9%; Pred. No. 7e+03;
 Matches 3; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtata 7
 : : : : : :
 Db 6 UUUGAUA 12

RESULT 14
 US-08-334-847-8
 ; Sequence 8, Application US/08334847
 ; Patent No. 5693532
 ; GENERAL INFORMATION:
 ; APPLICANT: McSwiggen, James
 ; APPLICANT: Draper, Kenneth
 ; APPLICANT: Pavco, Pam
 ; APPLICANT: Woolf, Tod
 ; TITLE OF INVENTION: METHOD AND REAGENT FOR
 ; TITLE OF INVENTION: INHIBITING RESPIRATORY
 ; TITLE OF INVENTION: SYNCYTIAL VIRUS
 ; NUMBER OF SEQUENCES: 909
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Lyon & Lyon
 ; STREET: 633 West Fifth Street
 ; CITY: Los Angeles
 ; STATE: California
 ; COUNTRY: U.S.A.
 ; ZIP: 90071-2066
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: IBM P.C. DOS 5.0
 ; SOFTWARE: Word Perfect 5.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/334,847
 ; FILING DATE: No. 5693532ember 4, 1994
 ; PRIOR APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Warburg, Richard J.
 ; REGISTRATION NUMBER: 32,327
 ; REFERENCE/DOCKET NUMBER: 209/032
 ; TELEPHONE: (213) 489-1600
 ; TELEFAX: (213) 955-0440
 ; TELEX: 67-3510
 ; INFORMATION FOR SEQ ID NO: 8:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 15 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-08-334-847-8

Query Match 100.0%; Score 7; DB 1; Length 15;
 Best Local Similarity 42.9%; Pred. No. 7e+03;

Matches 3; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ttgtata 7
 : : : : : :
 Db 7 UUUGAUA 13

RESULT 15
 US-08-334-847-9
 ; Sequence 9, Application US/08334847
 ; Patent No. 5693532
 ; GENERAL INFORMATION:
 ; APPLICANT: McSwiggen, James
 ; APPLICANT: Draper, Kenneth
 ; APPLICANT: Pavco, Pam
 ; APPLICANT: Woolf, Tod
 ; TITLE OF INVENTION: METHOD AND REAGENT FOR
 ; TITLE OF INVENTION: INHIBITING RESPIRATORY
 ; TITLE OF INVENTION: SYNCYTIAL VIRUS
 ; NUMBER OF SEQUENCES: 909
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Lyon & Lyon
 ; STREET: 633 West Fifth Street
 ; CITY: Los Angeles
 ; STATE: California
 ; COUNTRY: U.S.A.
 ; ZIP: 90071-2066
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: IBM P.C. DOS 5.0
 ; SOFTWARE: Word Perfect 5.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/334,847
 ; FILING DATE: No. 5693532ember 4, 1994
 ; PRIOR APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Warburg, Richard J.
 ; REGISTRATION NUMBER: 32,327
 ; REFERENCE/DOCKET NUMBER: 209/032
 ; TELEPHONE: (213) 489-1600
 ; TELEFAX: (213) 955-0440
 ; TELEX: 67-3510
 ; INFORMATION FOR SEQ ID NO: 9:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 15 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-08-334-847-9

Query Match 100.0%; Score 7; DB 1; Length 15;
 Best Local Similarity 42.9%; Pred. No. 7e+03;
 Matches 3; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtata 7
 : : : : : :
 Db 6 UUUGAUA 12

Search completed: March 27, 2001, 08:19:22
 Job time: 5149 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 27, 2001, 07:38:21 ; Search time 2517.78 Seconds
(without alignments)
19.482 Million cell updates/sec

Title: US-09-380-826A-5
Perfect score: 7
Sequence: 1 ttgata 7

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues 15983484

Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: gb_estl1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: em_estba:*
42: em_estfun:*
43: em_esthum1:*

44: em_esthum2:*
45: em_esthum3:*
46: em_esthum4:*
47: em_esthum5:*
48: em_esthum6:*
49: em_esthum7:*
50: em_esthum8:*
51: em_esthum9:*
52: em_esthum10:*
53: em_esthum11:*
54: em_esthum12:*
55: em_esthum13:*
56: em_esthum14:*
57: em_esthum15:*
58: em_esthum16:*
59: em_esthum17:*
60: em_esthum18:*
61: em_esthum19:*
62: em_esthum20:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estov1:*
68: em_estov2:*
69: em_estpl1:*
70: em_estpl2:*
71: em_estpl3:*
72: em_estpl4:*
73: em_estpl5:*
74: em_estrol1:*
75: em_estrol2:*
76: em_estrol3:*
77: em_estrol4:*
78: em_estrol5:*
79: em_estrol6:*
80: em_estrol7:*
81: em_estrol8:*
82: em_estrol9:*
83: em_estrol10:*
84: em_estrol11:*
85: em_estrol12:*
86: em_estrol13:*
87: gb_est41:*
88: gb_est42:*
89: gb_est43:*
90: gb_est44:*
91: gb_est45:*
92: gb_est46:*
93: gb_est47:*
94: gb_est48:*
95: gb_est49:*
96: gb_est50:*
97: gb_est51:*
98: gb_est52:*
99: gb_est53:*
100: gb_est54:*
101: gb_est55:*
102: gb_est56:*
103: gb_est57:*
104: gb_est67:*
105: gb_est68:*
106: gb_est69:*
107: gb_est70:*
108: gb_est71:*
109: gb_est72:*
110: gb_est73:*
111: gb_est74:*
112: em_esthum21:*
113: em_esthum22:*
114: em_esthum23:*
115: em_estom1:*
116: em_estom2:*

117: em_estp16:*
118: em_estp17:*
119: em_estp18:*
120: em_estp19:*
121: em_estp20:*
122: em_estp21:*
123: em_estp22:*
124: em_estp23:*
125: em_estp24:*
126: em_estp25:*
127: em_estp26:*
128: em_estp27:*
129: em_estp28:*
130: em_estp29:*
131: em_estp30:*
132: em_estp31:*
133: em_estp32:*
134: em_estp33:*
135: em_estp34:*
136: em_estp35:*
137: em_estp36:*
138: em_estp37:*
139: em_estp38:*
140: em_estp39:*
141: em_estp40:*
142: em_estp41:*
143: em_estp42:*
144: em_estp43:*
145: em_estp44:*
146: em_estp45:*
147: em_estp46:*
148: em_estp47:*
149: em_estp48:*
150: em_estp49:*
151: em_estp50:*
152: em_estp51:*
153: em_estp52:*
154: em_estp53:*
155: em_estp54:*
156: em_estp55:*
157: em_estp56:*
158: em_estp57:*
159: em_estp58:*
160: em_estp59:*
161: em_estp60:*
162: em_estp61:*
163: em_estp62:*
164: em_estp63:*
165: em_estp64:*
166: em_estp65:*
167: em_estp66:*
168: em_estp67:*
169: em_estp68:*
170: em_estp69:*
171: em_estp70:*
172: em_estp71:*
173: em_estp72:*
174: em_estp73:*
175: em_estp74:*
176: em_estp75:*
177: em_estp76:*
178: em_estp77:*
179: em_estp78:*
180: em_estp79:*
181: em_estp80:*
182: em_estp81:*
183: em_estp82:*
184: em_estp83:*
185: em_estp84:*
186: em_estp85:*
187: em_estp86:*
188: em_estp87:*
189: em_estp88:*

190: gb_gss25:*
191: gb_gss26:*
192: gb_gss27:*
193: gb_gss28:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	7	100.0	20	170	AZ333820	AZ333820 IM0045A01
C 2	7	100.0	22	170	AZ333820	AZ333820 IM0045A01
C 3	7	100.0	22	173	AZ502952	AZ502952 IM0034C07
C 4	7	100.0	23	171	AZ345476	AZ345476 IM0080P11
C 5	7	100.0	24	171	AZ379215	AZ379215 IM0134M14
C 6	7	100.0	24	173	AZ480650	AZ480650 IM0302P17
C 7	7	100.0	25	28	AU008929	AU008929 AU008929
C 8	7	100.0	25	172	AZ404619	AZ404619 IM0173L20
C 9	7	100.0	26	89	AW333097	AW333097 SI7C3 AGS
C 10	7	100.0	27	140	D18735	D18735 MUSGS01797
C 11	7	100.0	27	170	AZ338549	AZ338549 IM0052D24
C 12	7	100.0	27	170	AZ335603	AZ335603 IM0055E13
C 13	7	100.0	30	28	AU007442	AU007442 AU007442
C 14	7	100.0	30	137	BE911076	BE911076 601662175
C 15	7	100.0	30	139	C20899	C20899 HUMGS000497
C 16	7	100.0	31	16	A1140482	A1140482 ow82d12.s
C 17	7	100.0	31	27	A1973656	A1973656 sd07h11.y
C 18	7	100.0	32	28	AU009849	AU009849 AU009849
C 19	7	100.0	32	140	D18230	D18230 MUSGS00509
C 20	7	100.0	32	172	AZ447488	AZ447488 IM0244K07
C 21	7	100.0	33	145	R98634	R98634 yq69h06.r1
C 22	7	100.0	33	170	AZ308864	AZ308864 IM0012M13
C 23	7	100.0	33	172	AZ411435	AZ411435 IM0184C05
C 24	7	100.0	34	1	AA009774	AA009774 z104006.s
C 25	7	100.0	34	148	A0025283	A0025283 EP(3)3105
C 26	7	100.0	34	171	AZ387837	AZ387837 IM0147H19
C 27	7	100.0	34	173	AZ491530	AZ491530 IM0325H09
C 28	7	100.0	35	28	AU008005	AU008005 AU008005
C 29	7	100.0	36	23	A1630051	A1630051 ad00108.P
C 30	7	100.0	36	28	AU012250	AU012250 AU012250
C 31	7	100.0	36	28	AU012255	AU012255 AU012255
C 32	7	100.0	37	10	AA692338	AA692338 vt20f07.r
C 33	7	100.0	37	13	AA916682	AA916682 on14e10.s
C 34	7	100.0	37	172	AZ423769	AZ423769 IM0203G22
C 35	7	100.0	37	172	AZ429862	AZ429862 IM0214I05
C 36	7	100.0	37	173	AZ481957	AZ481957 IM0306F12
C 37	7	100.0	38	112	HS0002249	HS0002249 HOMO 88P1
C 38	7	100.0	38	137	BE886395	BE886395 601509756
C 39	7	100.0	38	139	C21092	C21092 HUMGS000260
C 40	7	100.0	39	141	H38033	H38033 YP58a09.s1
C 41	7	100.0	39	172	AZ402088	AZ402088 IM0169B16
C 42	7	100.0	39	172	AZ407162	AZ407162 IM0176A13
C 43	7	100.0	40	3	AA213154	AA213154 mw85c08.r
C 44	7	100.0	40	4	AA226145	AA226145 nc09f09.r
C 45	7	100.0	40	12	AA846437	AA846437 ai85c09.s

ALIGNMENTS

RESULT 1
AZ333820/c 20 bp DNA
LOCUS 1M0045A01R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
DEFINITION clone UUGC1M0045A01 R, DNA sequence.
ACCESSION AZ333820.1 GI:10378917
VERSION GSS.
KEYWORDS house mouse.
SOURCE

```

ORGANISM      Mus musculus
REFERENCE     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
1 (bases 1 to 20) Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS      Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
TITLE        Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL       Unpublished (2000)
COMMENT      Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0045 row: A column: 01
Seq primer: CACACAGGAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 20.
FEATURES     Location/Qualifiers
1..20
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="UGC1M0045A01"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT   10 a      4 c      3 g      3 t
ORIGIN
Query Match      100.0%; Score 7; DB 170; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ttgtata 7
|||||
Db 11 TTTGATA 5
RESULT 2
AZ329387/c      22 bp      DNA      GSS      29-SEP-2000
LOCUS          1M0053C07R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
DEFINITION     clone UUGC1M0053C07 R, DNA sequence.
ACCESSION      AZ329387
VERSION        AZ329387.1 GI:10390050
KEYWORDS       GSS.

SOURCE          house mouse.
Mus musculus
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 22)
REFERENCE      Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
TITLE        Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL       Unpublished (2000)
COMMENT      Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0053 row: C column: 07
Seq primer: CACACAGGAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 22.
FEATURES     Location/Qualifiers
1..22
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="UGC1M0053C07"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT      8 a      4 c      3 g      7 t
ORIGIN
Query Match      100.0%; Score 7; DB 170; Length 22;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ttgtata 7
|||||
Db 21 TTTGATA 15
RESULT 3
AZ502952/c      22 bp      DNA      GSS      05-OCT-2000
LOCUS          1M0342C04R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
DEFINITION     clone UUGC1M0342C04 R, DNA sequence.
ACCESSION      AZ502952
VERSION        AZ502952.1 GI:10684268
KEYWORDS       GSS.

```

VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
BASE COUNT
ORIGIN
Query Match
Best Local Similarity
Matches
QY
Db
RESULT
AZ345476/c
LOCUS
DEFINITION
ACCESSION

AZ345476.1 GI:10424713
GSS
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 23)
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.
and Wright, D., Weiss, R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0080 row: P column: 11
Seq primer: CGTTGTAACACGACGCCAGT
Class: plasmid ends
High quality sequence stop: 23.
Location/Qualifiers
1. .23
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0080P11"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PWD42 (gll4732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

100.0%; Score 7; DB 171; Length 23;
100.0%; Pred. No. 2.2e+05;
7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ttgtgata 7
|||||
23 TTTGATA 17

12 a 1 c 0 g 10 t

Query Match 100.0%; Score 7; DB 171; Length 23;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtgata 7
|||||
Db 23 TTTGATA 17

RESULT 5
AZ379215/c
LOCUS
DEFINITION
clone UUGC1M0134M14 F, DNA sequence.

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
BASE COUNT
ORIGIN
Query Match
Best Local Similarity
Matches
QY
Db
RESULT
AZ345476/c
LOCUS
DEFINITION
ACCESSION

AZ345476.1 GI:10424713
GSS
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 22)
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.
and Wright, D., Weiss, R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0342 row: C column: 04
Seq primer: CACACAGGAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 22.
Location/Qualifiers
1. .22
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0342C04"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PWD42 (gll4732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

100.0%; Score 7; DB 173; Length 22;
100.0%; Pred. No. 2.2e+05;
7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ttgtgata 7
|||||
9 TTTGATA 3

10 a 5 c 1 g .6 t

Query Match 100.0%; Score 7; DB 173; Length 22;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtgata 7
|||||
Db 9 TTTGATA 3

RESULT 4
AZ345476/c
LOCUS
DEFINITION
ACCESSION

AZ345476
1M0080P11F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC1M0080P11 F, DNA sequence.

ACCESSION AZ379215
 VERSION AZ379215.1 GI:10492915
 KEYWORDS GSS.
 SOURCE house musculus
 ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
 ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
 and Wright,D., Weiss,R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb

JOURNAL plasmid inserts
 COMMENT Unpublished (2000)
 Contact: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA

Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0134 row: M column: 14
 Seq primer: CGTTGTAAACGACGCCAGT
 Class: plasmid ends
 High quality sequence stop: 24.

FEATURES Location/Qualifiers

1..24
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC1M0134M14"
 /clone_lib="Mouse 10kb plasmid UUGC1M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
 /note="Vector: PWD42nv; Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of pWD42 (g114732114|gb|AF129072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

BASE COUNT 8 a 4 c 4 g 8 t
 ORIGIN

Query Match 100.0%; Score 7; DB 171; Length 24;
 Best Local Similarity 100.0%; Pred. No. 2.2e+05;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtgata 7
 |||||
 Db 12 TTTGATA 6

RESULT 6
 AZ480650/c 24 bp DNA GSS 04-OCT-2000
 LOCUS 1M0302P17F Mouse 10kb plasmid UUGC1M library Mus musculus genomic

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

REFERENCE
 AUTHORS

TITLE
 JOURNAL
 COMMENT

FEATURES Location/Qualifiers

1..24
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC1M0302P17"
 /clone_lib="Mouse 10kb plasmid UUGC1M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
 /note="Vector: PWD42nv; Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of pWD42 (g114732114|gb|AF129072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

BASE COUNT 11 a 4 c 1 g 8 t
 ORIGIN

Query Match 100.0%; Score 7; DB 173; Length 24;
 Best Local Similarity 100.0%; Pred. No. 2.2e+05;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtgata 7
 |||||
 Db 8 TTTGATA 2

RESULT 7
 AU008929/c 25 bp mRNA EST 31-JUL-1998
 LOCUS AU008929

DEFINITION AU008929 Schizosaccharomyces pombe late log phase cDNA
Schizosaccharomyces pombe cDNA clone spc04231, mRNA sequence.

ACCESSION AU008929
VERSION AU008929.1 GI:3345387
KEYWORDS EST.
SOURCE Schizosaccharomyces pombe

ORGANISM Schizosaccharomyces pombe
fission yeast.
Schizosaccharomycetes pombe
Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;
Schizosaccharomycetaceae; Schizosaccharomyces.

REFERENCE 1 (bases 1 to 25)
AUTHORS Morimyo,M. and Mita,K.
TITLE Identification of expressed sequence tags of Schizosaccharomyces pombe

JOURNAL Unpublished (1998)
COMMENT Contact: Mitsuoki Morimyo
Genome Research Group
National Institute of Radiological Sciences
9-1, Anagawa-4-chome, Inage-ku, Chiba, Chiba 263-8555, Japan
Email: morimyo@nirs.go.jp.
Location/Qualifiers

FEATURES
source
1..25
/organism="Schizosaccharomyces pombe"
/strain="972"
/db_xref="taxon:4896"
/clone="spc04231"
/clone_lib="Schizosaccharomyces pombe late log phase cDNA"
/sex="h minus"
/note="Vector: M13mp19; The cDNA library of Schizosaccharomyces pombe was prepared by cloning cDNA into the SmaI site of M13mp19 DNA and the direction of DNA sequences was not always from 5' to 3'. The cDNA data of Schizosaccharomyces pombe are available for searching on the World Wide Web. (URL, http://www.nirs.go.jp)"

BASE COUNT 16 a 3 c 2 g 4 t
ORIGIN

Query Match 100.0%; Score 7; DB 28; Length 25;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtgata 7
|||||
Db 11 TTTGATA 5

RESULT 8
AZ0404619 25 bp DNA GSS 03-OCT-2000
LOCUS IM0173L20F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0173L20 F, DNA sequence.

ACCESSION AZ0404619
VERSION AZ0404619.1 GI:10528632
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 25)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177

Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0173 row: L column: 20
Seq primer: CGTTGTAAACGACGCGCCACT
Class: plasmid ends
High quality sequence stop: 25.

FEATURES
source

1..25
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0173L20"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (gil4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 6 a 3 c 4 g 12 t
ORIGIN

Query Match 100.0%; Score 7; DB 172; Length 25;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtgata 7
|||||
Db 4 TTTGATA 10

RESULT 9
AW333097 26 bp mRNA EST 31-JAN-2000
LOCUS SL7C3 AGS-1 Pneumocystis carinii f. sp. carinii cDNA 3', mRNA sequence.
DEFINITION AW333097
ACCESSION AW333097.1 GI:6829454
VERSION AW333097
KEYWORDS EST.
SOURCE Pneumocystis carinii f. sp. carinii.
ORGANISM Pneumocystis carinii f. sp. carinii.
Pneumocystis carinii f. sp. carinii
Eukaryota; Fungi; Fungi incertae sedis; Pneumocystidaceae;

REFERENCE 1 (bases 1 to 26)
AUTHORS Smulian,A.G., Arnold,J., Weise,M., Wunderlich,J., Staben,C., Edman,J.C., Kovacs,J. and Cushion,M.
TITLE Expressed sequence tags from Pneumocystis carinii
JOURNAL Unpublished (2000)
COMMENT Contact: Staben C
School of Biological Sciences
University of Kentucky
101 Morgan Building, University of Kentucky, Lexington, KY 40506-0225, USA
Tel: 606 257 2161
Fax: 606 257 1717
Email: staben@pop.uky.edu.
Location/Qualifiers

```

source
1..26
/organism="Pneumocystis carinii f. sp. carinii"
/db_xref="taxon:38081"
/clone_lib="AGS-1"
/lab_host="E. coli"
/notes="Vector: Lambda ZAP II; Site_1: EcoRI; Site_2: XhoI;
P. carinii organisms (3x10e9) from a single rat (99-1-6,
sacrificed on 3/17/99) at Cincinnati VA facilities.
Trizol extracted RNA. Oligo dT priming, standard
conditions described by vendor, Stratagene. Further
details see www.uky.edu/project/Pneumocystis/"
BASE COUNT      13 a      2 c      1 g      10 t
ORIGIN
Query Match      100.0%; Score 7; DB 89; Length 26;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ttgtgata 7
Db 15 TTTGATA 21

RESULT 10
D18735
LOCUS      D18735      27 bp      mRNA      EST      12-DEC-1995
DEFINITION Mus musculus domesticus cdna clone
ACCESSION   D18735
VERSION     D18735.1 GI:1100704
KEYWORDS    EST.
SOURCE      western European house mouse.
ORGANISM   Mus musculus domesticus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1..(bases 1 to 27)
REFERENCE   1. (bases 1 to 27)
AUTHORS    Kawamoto,S., Okubo,K., Yoshii,J., Katsuki,M. and Matsubara,K.
TITLE      Analysis of gene expression in mouse embryogenesis by 3'-directed
cDNA sequencing
JOURNAL    Unpublished (1995)
COMMENT    Contact: Kawamoto,S., Okubo,K., Yoshii,J., Katsuki,M. and Matsubara
.K.
Institute for Cellular and Molecular Biology
Osaka University
3-1 Yamada-oka, Suita, Osaka 565, Japan.
FEATURES
source
1..27
/organism="Mus musculus domesticus"
/strain="C57BL/6J"
/db_xref="taxon:10092"
/clone_lib="md2066"
/tissue_type="decidual tissue (day 6.5-8.5 of gestation)"
BASE COUNT      11 a      4 c      4 g      8 t
ORIGIN
Query Match      100.0%; Score 7; DB 140; Length 27;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ttgtgata 7
Db 15 TTTGATA 21

RESULT 11
A2328549
LOCUS      A2328549      27 bp      DNA      GSS      29-SEP-2000
DEFINITION 1M0052D24F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC1M0052D24 F, DNA sequence.
ACCESSION   A2328549

```

```

VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
A2328549.1 GI:10388388
GSS.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1..(bases 1 to 27)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: dunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0052 row: D column: 24
Seq primer: CGTTGTAAACGACGCGCCAGT
Class: plasmid ends
High quality sequence stop: 27.
Location/Qualifiers
1..27
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="UUGC1M0052D24"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, P-"
/notes="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWD42 (gil4732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT      5 a      5 c      9 g      8 t
ORIGIN
Query Match      100.0%; Score 7; DB 170; Length 27;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ttgtgata 7
Db 4 TTTGATA 10

RESULT 12
A2335603/C
LOCUS      A2335603      27 bp      DNA      GSS      29-SEP-2000
DEFINITION 1M0065E13R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC1M0065E13 R, DNA sequence.

```

```

ACCESSION A2335603
VERSION A2335603.1 GI:10404082
KEYWORDS GSS.
SOURCE house musculus.
ORGANISM Mus musculus.

REFERENCE
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0065 row: E column: 13
Seq primer: CACACAGGAACACGCTATGACC
Class: plasmid ends
High quality sequence stop: 27.
FEATURES
source
Location/Qualifiers
1..27
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUC1M0065E13"
/clone_lib="Mouse 10kb plasmid UUC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/notes="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PWD42 (gil4732114[gblAF129072.1]), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT 9 a 8 c 1 g 9 t
ORIGIN

Query Match 100.0%; Score 7; DB 170; Length 27;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ttgtgata 7
Db 8 TTTGATA 2

RESULT 13
AU007442
LOCUS 30 bp mRNA EST
DEFINITION Schizosaccharomyces pombe late log phase cDNA
31-JUL-1998

ACCESSION AU007442
VERSION AU007442.1 GI:3343900
KEYWORDS EST.
SOURCE fission yeast.
ORGANISM Schizosaccharomyces pombe
Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;
Schizosaccharomycetaceae; Schizosaccharomyces.
REFERENCE
AUTHORS Morimyo,M. and Mita,K.
TITLE Identification of expressed sequence tags of Schizosaccharomyces
pombe
JOURNAL Unpublished (1998)
COMMENT Contact: Mitsuoki Morimyo
Genome Research Group
National Institute of Radiological Sciences
9-1, Anagawa-4-chome, Inage-ku, Chiba, Chiba 263-8555, Japan
Email: morimyo@nirs.go.jp
FEATURES
source
Location/Qualifiers
1..30
/organism="Schizosaccharomyces pombe"
/strain="972"
/db_xref="taxon:4896"
/clone="spc02033"
/clone_lib="Schizosaccharomyces pombe late log phase cDNA"
/sex="h minus"
/notes="vector: M13mpl9; The cDNA library of
Schizosaccharomyces pombe was prepared by cloning cDNA
into the SmaI site of M13mpl9 DNA and the direction of DNA
sequences was not always from 5' to 3'. The cDNA data of
Schizosaccharomyces pombe are available for searching on
the World Wide Web. (URL, http://www.nirs.go.jp)"
BASE COUNT 13 a 0 c 2 g 15 t
ORIGIN

Query Match 100.0%; Score 7; DB 28; Length 30;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ttgtgata 7
Db 21 TTTGATA 27

RESULT 14
BE911076
LOCUS 30 bp mRNA EST
DEFINITION 601662175p1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:3962130 5',
mRNA sequence.
ACCESSION BE911076
VERSION BE911076.1 GI:10408060
KEYWORDS EST.
SOURCE house musculus.
ORGANISM Mus musculus.
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 30)
NTH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9128 row: f column: 19
High quality sequence stop: 30.

```


FEATURES
source

Location/Qualifiers

1..30
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone_lib="IMAGE:3962130"
/clone_lib="NCI CGAP_Maml"
/tissue_type="tumor, biopsy sample"
/dev_stage="3 months, virgin"
/lab_host="DH10B"

/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"

BASE COUNT 12 a 1 c 4 g 13 t

ORIGIN

Query Match 100.0%; Score 7; DB 137; Length 30;

Best Local Similarity 100.0%; Pred. No. 2.2e+05;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtgata 7

|||||

Db 16 TTTGATA 22

RESULT 15

C20899/c

LOCUS

C20899 30 bp mRNA EST 23-OCT-1996
HUMGS0004973 Human adult (K.Okubo) Homo sapiens CDNA 3', mRNA
sequence.

ACCESSION

C20899 C20899.1 GI:1622009

VERSION

KEYWORDS EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 30)

Okubo, K.

BodyMap: human gene expression database

Unpublished (1995)

Contact: Okubo, K.

Institute for Molecular and Cellular Biol

Osaka University

1-3,Yamada-oka, Suita, Osaka Pref. 565, Japan

Tel: 06-877-5111(ex.3315)

Email: kousaku@imcb.osaka-u.ac.jp

Human Gene Signature. 3'-directed cDNA sequence. We are not

submitting the same cDNA sequence redundantly to DDBJ since 1993.

For the abundance information of clones with this sequence in this

library and as well as in other 3'-directed libraries, see

http://www.imcb.osaka-u.ac.jp/bodymap/. The sequences of the clones

represented by this GS sequences is also found there.

represented by this GS sequences is also found there.

represented by this GS sequences is also found there.

represented by this GS sequences is also found there.

represented by this GS sequences is also found there.

represented by this GS sequences is also found there.

represented by this GS sequences is also found there.

represented by this GS sequences is also found there.

represented by this GS sequences is also found there.

represented by this GS sequences is also found there.

represented by this GS sequences is also found there.

represented by this GS sequences is also found there.

represented by this GS sequences is also found there.

represented by this GS sequences is also found there.

represented by this GS sequences is also found there.

represented by this GS sequences is also found there.

represented by this GS sequences is also found there.

represented by this GS sequences is also found there.

represented by this GS sequences is also found there.

Search completed: March 27, 2001, 07:38:23
Job time: 4586 sec

FEATURES

source

Location/Qualifiers

1..30
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Human adult (K.Okubo)"
/dev_stage="adult"

BASE COUNT 14 a 6 c 5 g 5 t

ORIGIN

Query Match 100.0%; Score 7; DB 139; Length 30;

Best Local Similarity 100.0%; Pred. No. 2.2e+05;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtgata 7

|||||

Db 21 TTTGATA 15

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 27, 2001, 08:17:04 ; Search time 2286.42 Seconds
(without alignments)
49.243 Million cell updates/sec

Title: US-09-380-826A-6
Perfect score: 22
Sequence: 1 tgttgannnnnnnnnnttgata 22

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1118133 seqs, 2558875100 residues

Total number of hits satisfying chosen parameters: 2236266

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

- 1: gb_ba1.*
- 2: gb_ba2.*
- 3: gb_om.*
- 4: gb_ov.*
- 5: gb_ph.*
- 6: gb_pl1.*
- 7: gb_pl2.*
- 8: gb_pr1.*
- 9: gb_pr2.*
- 10: gb_pr3.*
- 11: gb_ro.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: em_fun.*
- 15: em_hum1.*
- 16: em_hum2.*
- 17: em_in.*
- 18: em_om.*
- 19: em_or.*
- 20: em_ov.*
- 21: em_pat.*
- 22: em_ph.*
- 23: em_pl.*
- 24: em_ro.*
- 25: em_sts.*
- 26: em_sy.*
- 27: em_un.*
- 28: em_v1.*
- 29: gb_ba3.*
- 30: gb_in1.*
- 31: gb_in2.*
- 32: gb_in3.*
- 33: gb_pl3.*
- 34: gb_pr4.*
- 35: em_ba1.*
- 36: em_ba2.*
- 37: em_htg1.*
- 38: em_htg2.*
- 39: em_htg3.*
- 40: em_htg4.*
- 41: em_htg5.*
- 42: em_htg6.*
- 43: em_htg7.*

- 44: em_htg8.*
- 45: em_htg9.*
- 46: em_htg10.*
- 47: em_hum3.*
- 48: em_hum4.*
- 49: em_hum5.*
- 50: em_hum6.*
- 51: gb_pr5.*
- 52: gb_pr6.*
- 53: gb_pr7.*
- 54: gb_htg1.*
- 55: gb_htg2.*
- 56: gb_htg3.*
- 57: gb_htg4.*
- 58: gb_htg5.*
- 59: gb_htg6.*
- 60: gb_htg7.*
- 61: gb_htg8.*
- 62: gb_htg9.*
- 63: gb_htg10.*
- 64: gb_htg11.*
- 65: gb_htg12.*
- 66: gb_htg13.*
- 67: gb_htg14.*
- 68: gb_htg15.*
- 69: gb_htg16.*
- 70: gb_htg17.*
- 71: gb_htg18.*
- 72: gb_htg19.*
- 73: gb_htg20.*
- 74: gb_htg21.*
- 75: gb_htg22.*
- 76: gb_htg23.*
- 77: gb_sts1.*
- 78: gb_sts2.*
- 79: gb_v11.*
- 80: gb_v12.*
- 81: gb_pat1.*
- 82: gb_pat2.*
- 83: em_htg0.*
- 84: gb_htg24.*
- 85: gb_pr8.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	14	63.6	239	53	HSSSG03	AF051649 Homo sapi
2	14	63.6	435	78	HUM4STS161	M95263 Human chrom
3	14	63.6	668	77	G50458	G50458 SHGC-79916
4	14	63.6	1015	80	SRSSRSV	D38547 SHC1 round
5	14	63.6	1257	3	BTBJ9225	AJ009225 Bos tauru
6	14	63.6	1257	52	HSAJ9229	AJ009229 Homo sapi
7	14	63.6	1258	3	BTBJ9224	AJ009224 Bos tauru
8	14	63.6	1276	3	BTBJ9223	AJ009223 Bos tauru
9	14	63.6	1287	53	HSU52077	U52077 Human marin
10	14	63.6	1481	2	LFU60594	U60594 Leptospira
11	14	63.6	2321	32	DDU66913	U66913 Dictyosteli
12	14	63.6	2683	29	SPGROELGN	X89236 S.pyogenes
13	14	63.6	3666	81	A45819	A45819 Sequence 13
14	14	63.6	3666	81	A45820	A45820 Sequence 14
15	14	63.6	3915	33	SCYJL046W	Z43321 S.cerevisia
16	14	63.6	4197	81	A45813	A45813 Sequence 7
17	14	63.6	4197	81	A45814	A45814 Sequence 8
18	14	63.6	4684	33	SCYJL045W	Z43320 S.cerevisia
19	14	63.6	5620	11	MMTHREC02	U36757 Mus musculu
20	14	63.6	10948	58	AC014767	AC014767 Drosophil
21	14	63.6	11949	51	AF000145	AF000145 Homo sapi

```

c 22 14 63.6 12437 1 AF002102 Ureaplasma
c 23 14 63.6 12973 57 AC013026 Drosophila
c 24 14 63.6 12978 30 AC006805 Caenorhab
c 25 14 63.6 28923 31 CEM03B6
c 26 14 63.6 37923 31 CELK07H8
c 27 14 63.6 42042 31 CEC34B4
c 28 14 63.6 53012 52 HSABLGR2
c 29 14 63.6 63156 65 AC026482 Homo sapi
c 30 14 63.6 63739 34 AF271897 Homo sapi
c 31 14 63.6 64064 71 AC083757 Homo sapi
c 32 14 63.6 64577 30 AE002828 Drosophila
c 33 14 63.6 64789 72 AC083839 Homo sapi
c 34 14 63.6 67987 58 AC016064 Homo sapi
c 35 14 63.6 68082 58 AC014851 Homo sapi
c 36 14 63.6 71842 66 AC036189 Homo sapi
c 37 14 63.6 74188 61 AC021315 Homo sapi
c 38 14 63.6 74562 52 HS25D22 Homo sapi
c 39 14 63.6 78877 9 AC008408 Homo sapi
c 40 14 63.6 85127 64 AC025978 Homo sapi
c 41 14 63.6 86113 34 AF285442 Homo sapi
c 42 14 63.6 88735 72 AF165178 Homo sapi
c 43 14 63.6 88800 51 HS127M18 Homo sapi
c 44 14 63.6 88958 9 AC008967 Homo sapi
c 45 14 63.6 91470 6 AC011438 Genomic s

```

ALIGNMENTS

```

RESULT 1
HSSG03 239 bp DNA 14-MAR-1999
LOCUS Homo sapiens squalene synthase gene, intron 2, 5' end.
DEFINITION AF051649
ACCESSION AF051649.1 GI:4415976
VERSION 3 of 14
KEYWORDS human.
SEGMENT
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
TITLE 1 (bases 1 to 239)
Jiang, G., McKenzle, T.L., Conrad, D.G. and Shechter, I.
Transcriptional regulation by lovastatin and 25-hydroxycholesterol
in HepG2 cells and molecular cloning and expression of the cDNA for
the human hepatic squalene synthase
J. Biol. Chem. 268 (17), 12818-12824 (1993)
JOURNAL 93266128
MEDLINE
REFERENCE 2 (bases 1 to 239)
Guan, G., Dai, P.H. and Shechter, I.
Direct Submission
TITLE Submitted (02-MAR-1998) Biochemistry and Molecular Biology,
JOURNAL Unifomed Services University of the Health Sciences, 4301 Jones
Bridge Road, Bethesda, MD 20814-4799, USA
FEATURES
Source
Location/Qualifiers
1..239
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="8"
/map="8p22-p23.1"
33..>239
/gene="squalene synthase"
/number=2
BASE COUNT 43 a 61 c 70 g 65 t
ORIGIN
|||||
1 tggtggannnnnnnttggata 22
|||||

```

```

Query Match 63.6%; Score 14; DB 53; Length 239;
Best Local Similarity 63.6%; Pred. No. 1.3e+03;
Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 tggtggannnnnnnttggata 22
|||||

```

```

Db 92 TGTTCGAAGCTACCTTTTGATA 113
RESULT 2
HUMASTS161
LOCUS HUMASTS161
DEFINITION Human chromosome 4 sequence-tagged site STS4-161, sequence tagged
site.
ACCESSION M95263
VERSION M95263.1 GI:177262
KEYWORDS STS; human chromosome 4; sequence tagged site.
SOURCE Homo sapiens, clone C4-177 from Los Alamos National Laboratory
chromosome 4 cosmid library, plate 4-1-2R, pos3C; vector SuperCos 1
(Stratagene).
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
REFERENCE Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
AUTHORS Gould, R.D., disibio, G., Xu, H., Lang, D.B., Dadgar, J., Magrane, G.,
Dugaiczky, A., Smith, K.A., Cox, D.R., Masters, S.B. and Myers, R.M.
TITLE The development of sequence-tagged sites for human chromosome 4
JOURNAL Hum. Mol. Genet. 2 (8), 1271-1288 (1993)
MEDLINE 94004872
COMMENT PCR components: 25 ng of human genomic DNA, 10 pmol of each
oligonucleotide, 200 micro-M dNTPs, 0.25 U Taq polymerase (Cetus)
in 10 micro-l of 50 mM KCl-20 mM Tris-HCl, pH 8.3 (at room temp),
2.5 mM MgCl-2. Initial denaturation at 94degC for 1.5 min, then 30
cycles of 94degC for 15 sec, 62degC for 23 sec, and 72degC for 30
sec, followed by a final extension at 72degC for 3.5 min, using a
perkin-Elmer 9600 thermocycler. PCR-amplified product size 279 bp.
Sequence submitted by:
Human Genome Mapping Center
Box 0925
University of California San Francisco
San Francisco, CA 94143-0925 USA
Phone: (415) 502-1612 Fax: (415) 476-8391
e-mail: hgmaprobes@cgl.ucsf.edu.
FEATURES
Source
1..435
/organism="Homo sapiens"
/db_xref="taxon:9606"
primer_bind complement(62..84)
primer_bind 316..340
BASE COUNT 128 a 64 c 75 g 162 t 6 others
ORIGIN
Query Match 63.6%; Score 14; DB 78; Length 435;
Best Local Similarity 63.6%; Pred. No. 1.2e+03;
Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 tggtggannnnnnnttggata 22
|||||
Db 262 TGTTCGAATTCTGCTTTTGATA 283
RESULT 3
G50458
LOCUS SHGC-79916 Human Homo sapiens STS genomic, sequence tagged site.
DEFINITION G50458
ACCESSION G50458
VERSION G50458.1 GI:5221635
KEYWORDS STS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
REFERENCE Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
AUTHORS 1 (bases 1 to 668)
Olivier, M. and Cox, D.R.
TITLE Unpublished, Olivier, M., Cox, D.R. (2000)
JOURNAL Unpublished (2000)
COMMENT Contact: Michael Olivier, David R. Cox

```

Stanford Human Genome Center
Stanford University School of Medicine
4005 Miranda Ave. 2nd Fl., Palo Alto, CA 94025, USA
Tel: (650) 320-5800
Fax: (650) 320-5801
Email: olivier@shgc.stanford.edu
Primer A: GCACATTAGCTTTGGATTGCTT
Primer B: TGCCTTGTGGTTTACCAATTTT
STS size: 278
PCR Profile:

Initial incubation: 95 degrees C for 10 minutes
Denaturation: 94 degrees C for 30 seconds
Annealing: 60 degrees C for 30 seconds
Polymerization: 72 degrees C for 23 seconds
PCR Cycles: 30
Thermal Cycler: Perkin Elmer 9700

Protocol:
Template: 25 ng
Primer: each 1 uM
dNTPs: each 200 uM
AmpliTaq Gold Polymerase: 0.07 units/ul
Total Vol: 5 ul

Buffer:
MgCl2: 2.5 mM
KCl: 50 mM
Tris-HCl: 10 mM
pH: 8.3

BAC ends sequenced at TIGR from the RPC111 BAC library. Designed and developed at the Stanford Human Genome Center.

FEATURES

source
1. .668
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="8"
/clone_lib="Human"

STS
primer_bind 31. .308
primer_bind 31. .53
BASE COUNT 230 a 104 c 113 g 221 t

ORIGIN

Query Match 63.6%; Score 14; DB 77; Length 668;
Best Local Similarity 63.6%; Pred. No. 1.2e+03;
Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 tgttgannnnnnnttgata 22
|||||

Db 543 TGTGGAAAATTACCTTGATA 564

RESULT 4

SRSSRSV/c SRSSRSV 1015 bp RNA VRL 08-FEB-1999
LOCUS Small round structured virus genomic RNA, 3' terminal sequence
DEFINITION containing ORF2 and ORF3.
ACCESSION D38547
VERSION D38547.1 GI:560054
KEYWORDS

SOURCE Small round structured virus (isolate:patient; SRSV-CHIBA-407/87/J)
cdna to genomic RNA.
ORGANISM small round structured virus
Viruses; ssRNA positive-strand viruses, no DNA stage;
Caliciviridae.
1 (bases 1 to 1015)
Utawaga,E.T.
Direct Submission
Submitted (17-OCT-1994) to the DDBJ/EMBL/GenBank databases. Etsuko
T. Utawaga, National Institute of Health; 1-23-1 Toyama,
Shinjuku-ku, Tokyo 162, Japan (Tel:03-5285-1111(ex.2562),
Fax:03-5285-1177)

REFERENCE

AUTHORS
TITLE
JOURNAL
BASE COUNT 388 a 263 c 263 g 343 t

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
FEATURES

2 (bases 1 to 1015)
Utawaga,E.T., Takeda,N., Inouye,S., Kasuga,K. and Yamazaki,S.
3'-terminal sequence of a small round structured virus (SRSV) in
Japan
Arch. Virol. 135 (1-2), 185-192 (1994)
94256839
Location/Qualifiers
1. .1015
/organism="small round structured virus"
/isolate="patient; SRSV-CHIBA-407/87/J"
/db_xref="taxon:37141"
/note="SRSVs are currently recognised as a potential human
calicivirus."
<1. .298
/note="ORF2"
/codon_start=2
/protein_id="BAA07549.1"
/db_xref="GI:1232112"
/translation="VPCLLPQBYITHFTISEQAPIQGEAALLHYVDPDTRNRLGEFKLY
PGYLITCVPNSSSTGPGQLPLDGVFVFSWVSRYQLKPVGTAGPARGLRVRR"
298. .924
/note="ORF3"
/codon_start=1
/protein_id="BAA07550.1"
/db_xref="GI:1232113"

BASE COUNT

ORIGIN

Query Match 63.6%; Score 14; DB 80; Length 1015;
Best Local Similarity 63.6%; Pred. No. 1.1e+03;
Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 tgttgannnnnnnttgata 22
|||||

Db 481 TGTGGACATTTCCATTTGATA 460

RESULT 5

BTJ9225 1257 bp DNA MAM 11-JUL-2000
LOCUS Bos taurus mariner related transposon Hsmar1, clone btmlc4.
DEFINITION
ACCESSION AJ009225
VERSION AJ009225.1 GI:9187451
KEYWORDS transposon.
SOURCE Bos taurus.
ORGANISM Bos taurus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 1257)
Demattell,M.V., Auge-Gouillou,C., Pollet,N., Meunier-Rotival,M. and
Bigot,Y.
Features of the mammal mar1 transposons in the human, sheep, cow
and mouse genomes and implications for their evolution
Mamm. Genome in press
2 (bases 1 to 1257)
Bigot,Y.
Direct Submission
Submitted (04-JUN-1998) Bigot Y., Institut de Recherche sur la
Biologie de l'Insectes, UPRESA CNRS 6035, Facult des Sciences, Parc
Grandmont, 37200 Tours, FRANCE
Location/Qualifiers
1. .1257
/organism="Bos taurus"
/transposon="Hsmar1"
/db_xref="taxon:9913"
/clone="btmlc4"

BASE COUNT

ORIGIN

Wed Mar 28 14:03:23 2001

```

AUTHORS      Bigot, Y.
TITLE        Direct Submission
JOURNAL      Submitted (04-JUN-1998) Bigot Y., Institut de Recherche sur la
              Biologie de l'Insectes, UPRESA CNRS 6035, Facult des Sciences, Parc
              Grandmont, 37200 Tours, FRANCE
FEATURES     Location/Qualifiers
              source
              1..1258
              /organism="Bos taurus"
              /transposon="Hsmar1"
              /db_xref="taxon:9913"
              /clone="btmlec3"
BASE COUNT   386 a 262 c 266 g 344 t
ORIGIN
Query Match  63.6%; Score 14; DB 3; Length 1257;
Best Local Similarity 63.6%; Pred. No. 1.1e+03;
Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 ttttgannnnnnnttgata 22
    |||||||
Db 36 TGTGGAAATTTGCCATTGATA 57

RESULT 6
HSAJ9229      1257 bp DNA PRI 11-JUL-2000
LOCUS        Homo sapiens mariner related transposon Hsmar1, clone mlehsc3.
DEFINITION   AJ009229
ACCESSION    AJ009229.1 GI:9187522
VERSION      transposon.
KEYWORDS     human.
SOURCE       Homo sapiens
ORGANISM     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE    1 (bases 1 to 1257)
AUTHORS      Demattei,M.V., Auge-Gouillou,C., Pollet,N., Meunier-Rotival,M. and
              Bigot,Y.
TITLE        Features of the mammal marl transposons in the human, sheep, cow
              and mouse genomes and implications for their evolution
JOURNAL      Mamm. Genome In press
REFERENCE    2 (bases 1 to 1257)
AUTHORS      Bigot,Y
TITLE        Direct Submission
JOURNAL      Submitted (04-JUN-1998) Bigot Y., Institut de Recherche sur la
              Biologie de l'Insectes, UPRESA CNRS 6035, Facult des Sciences, Parc
              Grandmont, 37200 Tours, FRANCE
FEATURES     Location/Qualifiers
              source
              1..1257
              /organism="Homo sapiens"
              /transposon="Hsmar1"
              /db_xref="taxon:9606"
              /clone="mlehsc3"
BASE COUNT   381 a 263 c 264 g 349 t
ORIGIN
Query Match  63.6%; Score 14; DB 52; Length 1257;
Best Local Similarity 63.6%; Pred. No. 1.1e+03;
Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 ttttgannnnnnnttgata 22
    |||||||
Db 36 TGTGGAAATTTGCCATTGATA 57

RESULT 7
HSAJ9224      1258 bp DNA MAM 11-JUL-2000
LOCUS        Bos taurus mariner related transposon Hsmar1, clone btmlec3.
DEFINITION   AJ009224
ACCESSION    AJ009224.1 GI:9187450
VERSION      transposon.
KEYWORDS     Bos taurus.
SOURCE       Bos taurus
ORGANISM     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
              Bovidae; Bovinae; Bos.
REFERENCE    1 (bases 1 to 1258)
AUTHORS      Demattei,M.V., Auge-Gouillou,C., Pollet,N., Meunier-Rotival,M. and
              Bigot,Y.
TITLE        Features of the mammal marl transposons in the human, sheep, cow
              and mouse genomes and implications for their evolution
JOURNAL      Mamm. Genome In press
REFERENCE    2 (bases 1 to 1258)

AUTHORS      Bigot, Y.
TITLE        Direct Submission
JOURNAL      Submitted (04-JUN-1998) Bigot Y., Institut de Recherche sur la
              Biologie de l'Insectes, UPRESA CNRS 6035, Facult des Sciences, Parc
              Grandmont, 37200 Tours, FRANCE
FEATURES     Location/Qualifiers
              source
              1..1276
              /organism="Bos taurus"
              /transposon="Hsmar1"
              /db_xref="taxon:9913"
              /clone="btmlec2"
BASE COUNT   388 a 261 c 248 g 379 t
ORIGIN
Query Match  63.6%; Score 14; DB 3; Length 1276;
Best Local Similarity 63.6%; Pred. No. 1.1e+03;
Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 ttttgannnnnnnttgata 22
    |||||||
Db 37 TGTGGAAATTTGCCATTGATA 58

RESULT 9
HSAJ92077     1287 bp DNA PRI 26-JAN-1998
LOCUS        Human mariner1 transposase gene, complete consensus sequence.
DEFINITION   HSU52077
ACCESSION    U52077
VERSION      U52077.1 GI:1263080

```

KEYWORDS human.
SOURCE SOURCE
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 1287)
JOURNAL Roberton,H.M. and Zumpano,K.L.
MEDLINE Molecular evolution of an ancient mariner transposon, Hsmar1, in
AUTHORS the human genome
TITLE Gene 205 (1-2), 203-217 (1997)
JOURNAL 98121293
MEDLINE 2 (bases 1 to 1287)
AUTHORS Roberton,H.M.
TITLE Direct Submission
JOURNAL Submitted (21-MAR-1996) Hugh M. Robertson, Entomology, University
of Illinois at Urbana-Champaign, 505 S. Goodwin, Urbana, IL 61801.
USA

FEATURES
source Location/Qualifiers
1..1287
/organism="Homo sapiens"
/db_xref="taxon:9606"
/transposon="Hsmar1"
/note="consensus sequence based on 20 unique sequences"
repeat_region 1..30
/rpt_type=inv
CDS 179..1210
/codon_start=1
/product="mariner transposase"
/protein_id="AAC52010.1"
/db_xref="GI:1263081"
/translation="MEMMLDKKQIRAIIFEFKMGKKAETTRINNNAFGPGTANER
VQWFKFKCGDELSGRPSVDNDQLRAII EADPLTTTREVABELNVHST
VHLQIGKGVKKLQWPHFELSENOKNRREVSSLLRNNPEFLDIRVTCDEKVI
YDNRRAQWLDREAPKHPKPNLHOKKVMTVMWSAAGLIHYSLNPGETTITSEK
AQQIDEMHKRLQRPALYNRKGPILLHDNARPHVAQPTLOKLNELGYEVLPHPPYS
AQQIDEMHKRLQRPALYNRKGPILLHDNARPHVAQPTLOKLNELGYEVLPHPPYS
DLSGDSYFD"
repeat_region 1256..1287
/rpt_type=inv
BASE COUNT 381 a 284 c 278 g 344 t
ORIGIN

Query Match 63.6%; Score 14; DB 53; Length 1287;
Best Local Similarity 63.6%; Pred. No. 1.1e+03;
Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0

QY 1 tgttgagannnnnnnttgata 22
|||||||
Db 36 TGTGGGAATTTGCCGTTTGATA 57

RESULT 10
LFU60594
LOCUS LFU60594 1481 bp DNA BCT 10-SEP-1998
DEFINITION Leptospira fainei 16S ribosomal RNA gene, partial sequence.
ACCESSION U60594
VERSION U60594.1 GI:1408219
KEYWORDS
SOURCE Leptospira fainei.
ORGANISM Leptospira fainei
REFERENCE Bacteria; Spirochaetales; Leptospiraceae; Leptospira.
AUTHORS 1 (bases 1 to 1481)
Perolat,P., Chappel,R.J., Adler,B., Baranton,G., Bulach,D.M.,
Billinghurst,M.L., Letocart,M., Merien,F. and Serrano,M.S.
TITLE Leptospira fainei sp. nov., isolated from pigs in Australia
JOURNAL Int. J. Syst. Bacteriol. 48 Pt 3, 851-858 (1998)
MEDLINE 98404550
REFERENCE 2 (bases 1 to 1481)
AUTHORS Adler,B., Chappel,R.J., Baranton,G., Bulach,D.M.,
Billinghurst,M.L., Letocart,M., Merien,F., Serrano,M.S. and
Perolat,P.
TITLE Direct Submission

JOURNAL Submitted (12-JUN-1996) Microbiology, Monash University, Wellington Rd., Clayton, VIC 3168, Australia

FEATURES

source 1. .1481
 /organism="Leptospira fainei"
 /strain="Hurstbridge"
 /db_xref="taxon:48782"
 <1. .>1481

rRNA
 /product="16S ribosomal RNA"
 391 a 335 c 439 g 314 t 2 others

BASE COUNT
 ORIGIN

Query Match 63.6%; Score 14; DB 2; Length 1481;
 Best Local Similarity 63.6%; Pred. No. 1.le+03;
 Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 tgttgagannnnnnnttgata 22
 ||||| |||||

Db 155 TGTGGATCACAAAGATTGTGATA 176

RESULT 11

DDU66913/c

LOCUS DDU66913 2321 bp DNA INV 05-SEP-1996

DEFINITION Dictyostellium discoideum ORF DG1040 gene, partial cds.

ACCESSION U66913

VERSION U66913.1 GI:1519537

KEYWORDS Dictyostellium discoideum.

SOURCE Dictyostellium discoideum.

ORGANISM Dictyostellium discoideum

REFERENCE 1 (bases 1 to 2321)
 Loomis, W.F.
 Dictyostellium discoideum
 Eukaryota; Dictyostellida; Dictyostellium.

AUTHORS Direct Submission

TITLE Submitted (15-AUG-1996) Dept. of Biology 0322, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA

JOURNAL

FEATURES

source 1. .2321
 /organism="Dictyostellium discoideum"
 /strain="AX4"
 /db_xref="taxon:44689"
 /cell_line="AX4"
 join(370. .732,813. .>2321)
 /note="ORF DG1040"
 <370. .732
 /number=1
 join(370. .732,813. .>2321)
 /note="ORF DG1040; initially derived from a plasmid disrupted gene"
 /codon_start=1
 /protein_id="AAB07544.1"
 /db_xref="GI:1519538"
 /translation="MTKPSIEDLKKSEVNDISINNNNNINNNNNNNNNKKEE
 TSYTPDNINIFNFKLSDDSKSYKWTQLQVGNNEQTSNTISNLSFVIATQIK
 LISNLTKQSKYKSNSELSQITNGYGSQAEIYIFRCLVDSIDFKNTNTPTTPTTSA
 TSTTPTPTTPTTASSTAASSSSNQLKQLPFKDEFSLTKOPHFISVLCRAFDGVE
 VLNNEFLVHFSPLKLTPTSHEIMGLALSQSLDKVEYREQADKFLNLTLQLSOSNTE
 LPENLHLQLLKFKIGTQLQNNQQNNQQNNQQNNQQNNQQNNQQNNQQNNQQNNQ
 QQNNQYSEVITYKLSLSPNSLALSPLVNDQSSGGGGITNSKRLQDDNPSYVN
 SILSKYQSNVPYQIIISIGYSSCSASTLKDLTFLQKLTESDIAQIIIVSMAELTSN
 TNNNNNNENNDENKTKIPYTFNDLNDKETNTPTSSSPKQSESSSSSTTTT
 NNNNNNNNTNTPDENWTSIFVDVITKELYPTIDWDIVIREMDCPVNNNDORGLAFI
 LAYKKATQVDQQQSPFPFIDFILDNRVWNSLSQIQFLKIAIQSDFFQNSTKKRID
 "

exon 813. .>2321
 /number=2

misc_feature 2124
 /note="site of plasmid insertion in mutant organisms; mutants form crinkled, branching fingers; neither spores nor stalk cells were seen"

BASE COUNT 1055 a 338 c 210 g 718 t

ORIGIN

Query Match 63.6%; Score 14; DB 32; Length 2321;
Best Local Similarity 63.6%; Pred. No. 1.le+03;
Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 ttttgannnnnnnttgata 22
|||||
Db 2069 TGTGGATATAATTCTTGATA 2048

RESULT 12
LOCUS SPGROELGN 2683 bp DNA BCT 26-SEP-1997
DEFINITION S. Pyogenes DNA for groEL gene.
ACCESSION X89236
VERSION X89236.1 GI:2462691
KEYWORDS groEL gene; heat shock protein 60 (GroEL) like protein.
SOURCE Streptococcus pyogenes.
ORGANISM Streptococcus pyogenes
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

REFERENCE 1 (bases 1 to 2683)
AUTHORS Podbielski, A.
TITLE Direct Submission
JOURNAL Submitted (29-JUN-1995) A. Podbielski, Institut f Med
Mikrobiologie, an der Rhein Westf. Techn. Hochschule, Aachen,
Pauwelsstr Klinikum, 52057 Aachen, FRG
2 (bases 1 to 2683)
AUTHORS Pohl, B., Podbielski, A. and Zarges, I.
JOURNAL Unpublished
COMMENT Related sequences M81132, M84965.

FEATURES
Location/Qualifiers
1..2683
/organism="Streptococcus pyogenes"
/strain="serotype M49"
/isolate="CS101"
/db_xref="taxon:1314"
1..1446
/gene="groEL"
<1..1446
/gene="groEL"
/codon_start=1
/transl_table=11
/product="heat shock protein"
/protein_id="CAA61520.1"
/db_xref="GI:2462692"
/db_xref="SPTREMBL:O33733"

translation="DFENMGAKLYSEVASKTNDIAGDGTATVLTQAIIVHEGLKNV
TAGNPIGIRGIETATATAEALKAQAPVSGKEAIAQAAVSRSEKVGVEYISEAM
ERVNDGVITIEESRGMETELVEVGMQFDKGLYLSQYMTDNERKMWADLENPFILMD
KRVNIQDILPLEELVLTNRPLLIADVDGEALPTILNKRIGTFNVAVKAPFG
DRKAMLEDAITLGTGTITDGLGLDAMTALGOAKKITVDKDSIVVEGSGSE
ATANRALIKSOLETTTSDFDREKLEQLAKLGGAVIKVGAPTEALKEMKLRID
ALNAPRAVEGIVAGGCTALTIVTEKVALEGGDATGRNIVLRALKEEPVROIALN
AGYESVVDKLNKSPAGTGFGNAATGEWDMIKTGIIDPVKVTSLQALQNAASVASLIL
TTEAVVANKPEAPAPAMPAMPMDPGMMGGF"
1509..1564
repeat_region
BASE COUNT 808 a 490 c 593 g 792 t
ORIGIN
/rpt_type=INVERTED

Query Match 63.6%; Score 14; DB 29; Length 2683;
Best Local Similarity 63.6%; Pred. No. 1.le+03;
Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 ttttgannnnnnnttgata 22
|||||
Db 963 TGTGGAGCTCCTACTTTGATA 942

RESULT 13

A45819/c
LOCUS A45819 3666 bp DNA PAT 07-MAR-1997
DEFINITION Sequence 13 from Patent WO9519371.
ACCESSION A45819
VERSION A45819.1 GI:2300192
KEYWORDS
SOURCE unidentified.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 3666)
AUTHORS Deblaere, R. Y., Desomer, J. and Dhaese, P.
TITLE EXPRESSION OF SURFACE LAYER PROTEINS
JOURNAL Patent: WO 9519371-A 13 20-JUL-1995;
SOLVAY (BE)
FEATURES Location/Qualifiers
1..3666
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 1281 a 662 c 676 g 1047 t
ORIGIN

Query Match 63.6%; Score 14; DB 81; Length 3666;
Best Local Similarity 63.6%; Pred. No. 1.le+03;
Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
Qy 1 ttttgannnnnnnttgata 22
|||||
Db 2652 TCTTGAGCAACTGGTTTGATA 2631

RESULT 14
LOCUS A45820/c 3666 bp DNA PAT 07-MAR-1997
DEFINITION Sequence 14 from Patent WO9519371.
ACCESSION A45820
VERSION A45820.1 GI:2300193
KEYWORDS
SOURCE unidentified.
ORGANISM unclassified.

REFERENCE 1 (bases 1 to 3666)
AUTHORS Deblaere, R. Y., Desomer, J. and Dhaese, P.
TITLE EXPRESSION OF SURFACE LAYER PROTEINS
JOURNAL Patent: WO 9519371-A 14 20-JUL-1995;
SOLVAY (BE)

FEATURES Location/Qualifiers
1..3666
/organism="unidentified"
/db_xref="taxon:32644"
1..3666
/note="unnamed protein product; Protein sequence is in
conflict with the conceptual translation"
/codon_start=1
/protein_id="CAA02849.1"
/db_xref="GI:2300194"
/translation="AQVNDYKISGYAKEAVALVDOGVIOGDTNGNPNPLNTVTRAQ
AAEITFKALELEANGDFKDKAGANTYINIAAVVANGIFEGVSATFEPANKSLTRS
EAAKILVEAFGLEGEADLSEFADASQVAPKAKYLEIAVANGIFEGTDANKLNPNNSI
TRQDFALVKRRTVDKVEGETPEEAAFKAINNTTVEVTVEEFTNVAQNFKEGLEI
KNASVKQTNKKVWVLTTEAQTADKEYVLTLDGETIGGFGVAAVPTKVELVSSAVQY
KLGOEVKQAKVTVAEGOSKAGIPVTETPGNNNDGVVPTLTGALTNEEGIATYSYT
RYKEGTDEYATATGDRSKFSLYVFWGVDTLSVEEVTTCASVNGANGKTYKVTYKN
PKTGPEAKNTFNVGFDVNMNVTSKVNANVNGVKALQSLNSGALDRAAQITTSKGE
ATFTVSGTNAATPVVYDLHSTNNSTSNKYSASALQTTASKVTFALQAEYITELTR
ADNGEVAAGATNGREYKIVDKAGNLAKNEIVNFAFNEDKDRVISTNNAKFPDVT
DPDTAVYFTGKAKQISVTKDKGEATIGSVTVDYATPIAMIDINTSDAKQGDLD
EGEPKAVAPISYFOAPYLDGSAIKAYKKSDLNKAVTKFDGSETAVFAELVNQSGKKV
TGTSIKKATYTYTNGANDIKVNOQVISPNSYTYVTYTSSTGTVITPAKNLEVTS
VDGKTAVKVIATGIANTDGDYAFATAKEATATATATNPNNSYTGATOPNFADSC
SNSNSIWFAGNPVKYAGVSGTKYKFGANGNEVFEAGEALITQYATEQOKVTISY
NVGDGTVTFKVISAVNSESITAKPVPATTPAAPTGTALTTPAAGGLVDLLTATNTLS
ISLADADLNVSATTVDTATVSLKDSANNLSLTLTETGANTGVFATTVQAGTLLSLTA

GTLTYADAKNAGVAENITASVTLKTKTGAITSDDTFTQGVLPNSATAAEYTSKISA
 ADYTAEGGFTLINDNAGQVINLAGKGAQGVADAINATFAGTATVSGDKVIRKSA
 TTGVSEVEVTFSSVQVLNAVYNGKQDVVAGTAATKAFITITALSVEKVVIDGVEY
 TAAVACTAPTANTFVVESSAANTLASVADQAAANLAATIDTLNTADKFTASATGATITLT
 SNTVPVGTITTEPVTILK"

BASE COUNT 1281 a 662 c 676 g 1047 t
 ORIGIN

Query Match 63.6%; Score 14; DB 81; Length 3666;

Best Local Similarity 63.6%; Pred. No. 1e+03; Mismatches 0; Gaps 0; Indels 8; Length 3666;

Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0; Indels 8; Length 3666;

QY 1 tgttgannnnnnntttgata 22

DB 2652 TGTGGAGCAACTGGTTTGATA 2631

RESULT 15

SCYJL046W

LOCUS SCYJL046W 3915 bp DNA PLN 11-AUG-1997

DEFINITION S.cerevisiae chromosome X reading frame ORF YJL046w.

ACCESSION 249321 Y13136

VERSION 249321.1 GI:1008176

KEYWORDS

SOURCE baker's yeast.

ORGANISM

Saccharomyces cerevisiae

Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;

Saccharomycetaceae; Saccharomycetes.

1 (bases 1 to 3915)

Pohl, T.M. and Aljinovic, G.

Unpublished

2 (bases 1 to 3915)

MIPS.

Direct Submission

Submitted (25-SEP-1995) Data collected by MIPS on behalf of the

European yeast chromosome X sequencing project. MIPS at the

Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152

Martinsried, FRG; E-mail: Mewes@mips.emblnet.org

Location/Qualifiers

1..3915

/organism="Saccharomyces cerevisiae"

/db_xref="taxon:4932"

/chromosome="X"

152..1507

/note="ORF YJL046w"

/codon_start=1

/protein_id="CAA89337.1"

/db_xref="GI:1008177"

/db_xref="SWISS-PROT:P47051"

/translation="MRSASSYLIRNIEHPKISELTYVEFLMNTKEENTRLFFDVIIMS

MMLSNWAISPRVVGQRNLHCTTLFHLTRWAKDADDKYHDINSYNNFTPSNDNVS

ILDEGKSDYDTTKASMEEDISAFNKDLYNFNIGYAKQIMSASOLENIVRAKGRFV

IQSLSTSPYINLALENYFKNTPRAKRPDNCRLLLYINDRCVAVIGKQNLQWQVDLA

KLKSKNFELLRFSGGTVLHLDGNVNSYLTLSREKFTKFNKMTIKWLSINPELR

LDLNERGDIIDQDKISQAYKIAGKAYHATMLLNADLQFSGLLPESLPNNWEWE

SGSVISVKSKINVGILITPNQFIIVYSERFKRVGGEIPIYICDEFKSIINDEKDA

MNTLQSQWKVSPGKFSVKIKDKGLTIKVERGMIVDCDRNDLIGLEFKGFLENIDSY

T"

2136..2224

/gene="ty(GTA)JL - systematic name"

/note="tRNA-Tyr - common name; anticodon gene: GTA;

contains intron 14 nt"

2136..2224

/gene="ty(GTA)JL - systematic name"

complement(2445..2735)

/note="delta-remnant"

complement(2736..3067)

/note="solo delta"

3266..3337

/gene="tr(TCT)JL2 - systematic name"

/note="tRNA-Arg2 - common name; anticodon gene: TCT"

3266..3337

gene

/gene="tr(TCT)JL2 - systematic name"
 3348..3419

/gene="td(GTC)JL2 - systematic name"

/note="tRNA-Asp - common name; anticodon gene: GTC"

3348..3419

/gene="td(GTC)JL2 - systematic name"

BASE COUNT 1258 a 667 c 698 g 1292 t

ORIGIN

Query Match 63.6%; Score 14; DB 33; Length 3915;

Best Local Similarity 63.8%; Pred. No. 1e+03; Mismatches 0; Gaps 0; Indels 8; Length 3915;

Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0; Indels 8; Length 3915;

QY 1 tgttgannnnnnntttgata 22

DB 2401 TGTGGAAATCCATTTTGTGATA 2422

Search completed: March 27, 2001, 08:17:06

Job time: 5903 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 27, 2001, 08:22:59 ; Search time 207.51 Seconds
(without alignments)
39.827 Million cell updates/sec.

Title: US-09-380-826A-6

Perfect score: 22
Sequence: 1 ttgtggannnnnnnttgata 22

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgn2_2/gcgdata/geneseq/geneseq/NA1980.DAT:*
2: /cgn2_2/gcgdata/geneseq/geneseq/NA1981.DAT:*
3: /cgn2_2/gcgdata/geneseq/geneseq/NA1982.DAT:*
4: /cgn2_2/gcgdata/geneseq/geneseq/NA1983.DAT:*
5: /cgn2_2/gcgdata/geneseq/geneseq/NA1984.DAT:*
6: /cgn2_2/gcgdata/geneseq/geneseq/NA1985.DAT:*
7: /cgn2_2/gcgdata/geneseq/geneseq/NA1986.DAT:*
8: /cgn2_2/gcgdata/geneseq/geneseq/NA1987.DAT:*
9: /cgn2_2/gcgdata/geneseq/geneseq/NA1988.DAT:*
10: /cgn2_2/gcgdata/geneseq/geneseq/NA1989.DAT:*
11: /cgn2_2/gcgdata/geneseq/geneseq/NA1990.DAT:*
12: /cgn2_2/gcgdata/geneseq/geneseq/NA1991.DAT:*
13: /cgn2_2/gcgdata/geneseq/geneseq/NA1992.DAT:*
14: /cgn2_2/gcgdata/geneseq/geneseq/NA1993.DAT:*
15: /cgn2_2/gcgdata/geneseq/geneseq/NA1994.DAT:*
16: /cgn2_2/gcgdata/geneseq/geneseq/NA1995.DAT:*
17: /cgn2_2/gcgdata/geneseq/geneseq/NA1996.DAT:*
18: /cgn2_2/gcgdata/geneseq/geneseq/NA1997.DAT:*
19: /cgn2_2/gcgdata/geneseq/geneseq/NA1998.DAT:*
20: /cgn2_2/gcgdata/geneseq/geneseq/NA1999.DAT:*
21: /cgn2_2/gcgdata/geneseq/geneseq/NA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	14	63.6	22	19 V58901	Leptospira rRNA ge
2	14	63.6	22	19 V58901	L. fainei nucleoti
3	14	63.6	22	19 V58900	Leptospira nucleot
4	14	63.6	972	20 X99627	Nucleic acid seque
5	14	63.6	1477	19 V58896	L. fainei nucleoti
6	14	63.6	1661	20 X86155	DNA encoding a SLR
7	14	63.6	4197	16 Q99430	B. sphaericus str
8	13	59.1	396	20 X34752	Human encoding OMP-1
9	13	59.1	828	20 Z16735	Human gene express
10	13	59.1	1079	20 X20676	Polynucleotide seq
11	13	59.1	1482	19 V18099	Nucleotide sequenc
12	13	59.1	1786	18 T60350	MAP kinase #2 codi

13	13	59.1	2190	20 X07102	Staphylococcus aur
14	13	59.1	2508	18 T67197	zebrafish retinoid
15	13	59.1	3519	21 Z94941	Human carbohydrate
16	13	59.1	3519	21 Z94948	Human carbohydrate
17	13	59.1	4012	18 V74358	Staphylococcus aur
18	13	59.1	5253	21 A26868	Essential Staphylo
19	13	59.1	9052	18 V74396	Staphylococcus aur
20	13	59.1	10723	18 T49304	CDNA encoding poly
21	13	59.1	10723	18 T49303	CDNA sequence enco
22	13	59.1	11802	18 V74381	Staphylococcus aur
23	13	59.1	15614	20 X12982	Enterococcus faeca
24	13	59.1	1230025	20 X91990	Nucleotide sequenc
25	13	59.1	1664976	19 V21209	Methanococcus jann
26	12.4	56.4	30	21 Z40032	Oligonucleotide fo
27	12.4	56.4	30	21 Z40033	Oligonucleotide fo
28	12.4	56.4	200	19 X12115	Human blaIIelic po
29	12.4	56.4	300	20 Z13959	Human gene express
30	12.4	56.4	300	20 Z12971	Human gene express
31	12.4	56.4	310	20 X20027	Enterococcus faeca
32	12.4	56.4	414	20 X20026	Enterococcus faeca
33	12.4	56.4	706	20 X20469	Human secreted pro
34	12.4	56.4	814	20 X35887	CDNA encoding a pr
35	12.4	56.4	877	18 V75289	Staphylococcus aur
36	12.4	56.4	915	19 V24146	Homo sapiens BARD1
37	12.4	56.4	963	21 A05842	Group B Streptococ
38	12.4	56.4	1034	18 T72785	Metastasis inducin
39	12.4	56.4	1131	18 V25036	H. pylori cellular
40	12.4	56.4	1152	18 V24747	H. pylori ORF hp4p
41	12.4	56.4	1248	21 A05803	Group B Streptococ
42	12.4	56.4	1249	19 X14031	H. pylori GHPO 886
43	12.4	56.4	1322	20 X13419	Enterococcus faeca
44	12.4	56.4	1353	18 X30790	Streptococcus pneu
45	12.4	56.4	1356	19 V04491	A. thaliana p-hydr

ALIGNMENTS

RESULT 1
V58901
ID V58901 standard; DNA; 22 BP.
XX
AC V58901;
XX
DT 20-JAN-1999 (first entry)
XX
DE Leptospira rRNA gene nucleotide sequence.
XX
KW Infection; pathogenic Leptospira; protective immunity; therapy;
diagnosis; ss.
XX
OS Leptospira sp.
XX
PN WO9840099-A1.
XX
PD 17-SEP-1998.
XX
PF 06-MAR-1998; 98WO-AU00145.
XX
PR 07-MAR-1997; 97AU-0005494.
XX
(AGRI-) AGRIC VICTORIA SERVICES PTY LTD.
(PIGR-) PIG RES & DEV CORP.
XX
Chappel RJ;
XX
WPI; 1998-520791/44.
XX
New isolated pathogenic Leptospira bacterium - useful for, e.g
developing products for conferring protective immunity, and for
prophylactic or therapeutic treatment
XX
Cialm 15; Page 72; 94pp; English.
PS

XX This sequence represents a Leptospira DNA sequence isolated from the
 CC pathogenic Leptospira (LS) bacterium of the invention. The bacterium
 CC belongs to serogroup Hurstbridge or serovar hurstbridge or the species
 CC L. fainei. The LS bacteria can be used for conferring protective
 CC immunity against pathogenic LS bacteria in humans or animals. The
 CC bacteria can also be used for prophylactic or therapeutic treatment of LS
 CC infections. The DNAs and antibodies may also be used for detection and
 CC diagnosis of past or present LS infection.
 XX
 SQ Sequence 22 BP; 7 A; 2 C; 5 G; 8 T; 0 other;

Query Match 63.6%; Score 14; DB 19; Length 22;
 Best Local Similarity 63.6%; Pred. No. 33;
 Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 ttgtggannnnnnntttgata 22
 |||||
 Db 1 ttgtggatcacagatttgata 22

RESULT 2

V58897
 ID V58897 standard; DNA; 22 BP.

XX
 AC V58897;

XX 20-JAN-1999 (first entry)

XX L. fainei nucleotide sequence.

XX Infection; pathogenic Leptospira; protective immunity; therapy;
 KW diagnosis; ss.

XX Leptospira fainei.

PN WO9840099-A1.

XX 17-SEP-1998.

XX 06-MAR-1998; 98WO-AU00145.

PR 07-MAR-1997; 97AU-0005494.

XX (AGRI-) AGRIC VICTORIA SERVICES PTY LTD.

PA (PIGR-) PIG RES & DEV CORP.

XX Chappel RJ;

XX WPI; 1998-520791/44.

XX New isolated pathogenic Leptospira bacterium - useful for, e.g
 PT developing products for conferring protective immunity, and for
 PT prophylactic or therapeutic treatment

PS Claim 15; Page 70; 94pp; English.

XX This sequence represents a Leptospira DNA sequence isolated from the
 CC pathogenic Leptospira (LS) bacterium of the invention. The bacterium
 CC belongs to serogroup Hurstbridge or serovar hurstbridge or the species
 CC L. fainei. The LS bacteria can be used for conferring protective
 CC immunity against pathogenic LS bacteria in humans or animals. The
 CC bacteria can also be used for prophylactic or therapeutic treatment of LS
 CC infections. The DNAs and antibodies may also be used for detection and
 CC diagnosis of past or present LS infection.
 XX

SQ Sequence 22 BP; 7 A; 2 C; 5 G; 8 T; 0 other;

Query Match 63.6%; Score 14; DB 19; Length 22;
 Best Local Similarity 63.6%; Pred. No. 33;
 Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 ttgtggannnnnnntttgata 22
 |||||
 Db 1 ttgtggatcacagatttgata 22

RESULT 3

V58900
 ID V58900 standard; DNA; 22 BP.

XX
 AC V58900;

XX 20-JAN-1999 (first entry)

XX Leptospira nucleotide sequence.

XX Infection; pathogenic Leptospira; protective immunity; therapy;
 KW diagnosis; ss.

XX Leptospira sp.

PN WO9840099-A1.

XX 17-SEP-1998.

XX 06-MAR-1998; 98WO-AU00145.

PR 07-MAR-1997; 97AU-0005494.

PA (AGRI-) AGRIC VICTORIA SERVICES PTY LTD.

PA (PIGR-) PIG RES & DEV CORP.

XX Chappel RJ;

XX WPI; 1998-520791/44.

XX New isolated pathogenic Leptospira bacterium - useful for, e.g
 PT developing products for conferring protective immunity, and for
 PT prophylactic or therapeutic treatment

PS Claim 15; Page 72; 94pp; English.

XX This sequence represents a Leptospira DNA sequence isolated from the
 CC pathogenic Leptospira (LS) bacterium of the invention. The bacterium
 CC belongs to serogroup Hurstbridge or serovar hurstbridge or the species
 CC L. fainei. The LS bacteria can be used for conferring protective
 CC immunity against pathogenic LS bacteria in humans or animals. The
 CC bacteria can also be used for prophylactic or therapeutic treatment of LS
 CC infections. The DNAs and antibodies may also be used for detection and
 CC diagnosis of past or present LS infection.
 XX

SQ Sequence 22 BP; 3 A; 0 C; 4 G; 7 T; 8 other;

Query Match 63.6%; Score 14; DB 19; Length 22;
 Best Local Similarity 100.0%; Pred. No. 33;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ttgtggannnnnnntttgata 22
 |||||
 Db 1 ttgtggannnnnnntttgata 22

RESULT 4

X99627/c

ID X99627 standard; DNA; 972 BP.

XX
 AC X99627;

XX 05-OCT-1999 (first entry)

XX Nucleic acid sequence from U. urealyticum.

KW Ureaplasma urealyticum; nucleic acid detection; infection; pathogen;
 KW human urogenital tract; pregnancy; neonatal disease; drug therapy;
 XX suppurative arthritis; ss.
 OS Ureaplasma urealyticum.
 PN WO9939007-A1.
 XX
 XX
 PD 05-AUG-1999.
 XX
 XX 29-JAN-1999; 99WO-US01972.
 XX
 XX 30-JAN-1998; 98US-0073189.
 XX (UABR-) UAB RES FOUND.
 PA
 XX Cassell GH, Chen EY, Glass JT, Glass JS, Heiner CR;
 PI Lefkowitz E;
 XX
 XX WPI; 1999-469343/39.
 DR
 XX Detection of Ureaplasma urealyticum using novel genes, probes and
 XX primers
 PT
 XX
 XX Claim 1; Page 82; 110pp; English.
 PS
 XX The present invention provides methods for the detection and diagnosis
 CC of Ureaplasma urealyticum infection. It provides novel genes (X99501-681)
 CC that can be used as a source of primers and probes for the detection and/
 CC or quantification of U. urealyticum in a biological sample. The probes
 CC that can be used in the method of the invention by forming target:probe
 CC complex is complementary to a region selected from one of the 181
 CC nucleic acid sequences (X99501-681). U. urealyticum is an opportunistic
 CC pathogen of the human urogenital tract that is a significant cause of
 CC adverse pregnancy outcome, neonatal disease, and suppurative arthritis.
 CC As the infections are commonly asymptomatic, it is important to have
 CC specific and sensitive methods for detecting their presence in a patient.
 CC Also, as the pathogen has no current antibiotic directed specifically
 CC against it, it would be advantageous to isolate and detect gene sequences
 CC which are unique to it, and utilise these as a basis for diagnosis of
 CC U. urealyticum infection as well as to develop new and improved drug
 CC therapies. The present invention provides such novel polynucleotide
 CC sequences (X99501-681).
 XX
 XX Sequence 972 BP; 309 A; 127 C; 98 G; 438 T; 0 other;
 SQ

Query Match 63.6%; Score 14; DB 20; Length 972;
 Best Local Similarity 63.6%; Pred. No. 53;
 Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 tgttggannnnnnntttgata 22
 |||||
 Db 931 TGTGGACTATTAGTTTGATA 910

RESULT 5
 V58896
 ID V58896 standard; DNA; 1477 BP.
 XX
 XX V58896;
 AC
 XX 20-JAN-1999 (first entry)
 DT
 XX L. fainei nucleotide sequence.
 DE
 XX Infection; pathogenic Leptospira; protective immunity; therapy;
 KW diagnosis; ss.
 XX Leptospira fainei.
 OS
 XX WO9840099-A1.
 PN
 XX

PD 17-SEP-1998.
 XX
 PF 06-MAR-1998; 98WO-AU00145.
 XX
 XX 07-MAR-1997; 97AU-0005494.
 XX
 XX (AGRI-) AGRIC VICTORIA SERVICES PTY LTD.
 PA (PIGR-) PIG RES & DEV CORP.
 XX
 XX Chappel RJ;
 PI
 XX WPI; 1998-520791/44.
 DR
 XX New isolated pathogenic Leptospira bacterium - useful for, e.g
 PT developing products for conferring protective immunity, and for
 PT prophylactic or therapeutic treatment
 PT
 XX Claim 15; Page 69-70; 94pp; English.
 PS
 XX This sequence represents a Leptospira DNA sequence isolated from the
 CC pathogenic Leptospira (LS) bacterium of the invention. The bacterium
 CC belongs to serogroup Hurstbridge or serovar Hurstbridge or the species
 CC L. fainei. The LS bacterium can be used for conferring protective
 CC immunity against pathogenic LS bacteria in humans or animals. The
 CC bacteria can also be used for prophylactic or therapeutic treatment of LS
 CC infections. The DNAs and antibodies may also be used for detection and
 CC diagnosis of past or present LS infection.
 XX
 XX Sequence 1477 BP; 390 A; 334 C; 439 G; 314 T; 0 other;
 SQ

Query Match 63.6%; Score 14; DB 19; Length 1477;
 Best Local Similarity 63.6%; Pred. No. 56;
 Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 tgttggannnnnnntttgata 22
 |||||
 Db 154 tgttgatcacagatttgata 175

RESULT 6
 X86155/C
 ID X86155 standard; DNA; 1661 BP.
 XX
 XX X86155;
 AC
 XX 22-SEP-1999 (first entry)
 DT
 XX DNA encoding a Streptococcus pyogenes heat shock protein (Hsp)60-2.
 DE
 XX Heat shock protein; Hsp60-2; immune response; immunological carrier;
 KW cancer control; tumour; sarcoma; cancer; gene therapy; ss.
 KW
 XX Streptococcus pyogenes.
 OS
 XX WO9935270-A1.
 PN
 XX 15-JUL-1999.
 PD
 XX 29-DEC-1998; 98WO-CA01203.
 PF
 XX 31-DEC-1997; 97US-0001737.
 PR
 XX (STRE-) STRESSGEN BIOTECHNOLOGIES CORP.
 PA
 XX Mizzen L, Wisniewski J;
 PI
 XX WPI; 1999-430397/36.
 DR
 DR P-PSDB; Y23904.
 XX
 XX New nucleic acid encoding heat shock protein-60 from Streptococcus,
 PT useful in vaccines, as carriers for other immunogens, as anticancer
 PT agents and for diagnosis

XX Claim 3; Fig 4A-B; 176pp; English.
 XX The present sequence encodes a heat shock protein, designated Hsp60-2.
 CC The protein, its fragments, variants and fusion proteins, are
 CC used to elicit or enhance an immune response against Streptococcus,
 CC and to elicit a similar response to a target antigen fused to the
 CC protein. Unlike other immunological carriers, Hsp60 proteins are not
 CC immunosuppressive so provide an increased response to any conjugated or
 CC fused antigen. Also, where used for cancer control, they lack the side
 CC effects associated with endotoxins. They can also be used to detect
 CC specific antibodies and in treatment or prevention of tumours
 CC (e.g. sarcoma or cancers of breast, ovary, prostate, lung, pancreas or
 CC liver). The Hsp60 polynucleotide is used for recombinant production
 CC of the protein, as a source of primers and probes for detecting
 CC streptococci in standard hybridization/amplification assays, and
 CC therapeutically in gene therapy vectors.
 XX Sequence 1661 BP; 512 A; 308 C; 407 G; 434 T; 0 other;

Query Match 63.6%; Score 14; DB 20; Length 1661;
 Best Local Similarity 63.6%; Pred. No. 57;
 Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 ttttgannnnnnnttgata 22
 |||||
 DB 1163 TGTGGAGCTCTACTTTGATA 1142

RESULT 7
 Q99430/C
 ID Q99430 standard; DNA; 4197 BP.
 XX AC Q99430;
 XX 22-DEC-1995 (first entry)
 XX B. sphaericus SLP gene.
 XX Surface layer protein; SLP; fusion protein; vaccine; antigen;
 KW surface expression; epitope; ds.
 XX Bacillus sphaericus.

Key Location/Qualifiers
 RBS 79..85
 FT /*tag= a
 FT 95..3853
 FT /*tag= b
 FT sig_peptide 95..184
 FT /*tag= c
 FT mat_peptide 185..3850
 FT /*tag= d
 XX W09519371-A2.
 XX 20-JUL-1995.
 XX 13-JAN-1995; 95WO-EP00147.
 XX 14-JAN-1994; 94GB-0000650.
 XX (SOLV) SOLVAY SA.

XX Deblaere RY, Desomer J, Dhaese P;
 XX WPI; 1995-263827/34.
 DR P-PSDB; R80530.
 XX Host cell expressing surface layer protein fusion protein - used for
 PT host presentation of antigens and vaccine prodn.
 XX

PS Disclosure; Fig.6; 95pp; English.

XX A probe based on the N-terminal sequence of B. sphaericus P-1 (LMG
 CC P-13855) surface layer protein was used to screen an HindIII-
 CC generated library to isolate the slp gene. Promoter regions
 CC of the gene are used in genetic constructs providing surface
 CC expression of heterologous proteins in P-1 hosts.
 XX Sequence 4197 BP; 1470 A; 731 C; 763 G; 1233 T; 0 other;

Query Match 63.6%; Score 14; DB 16; Length 4197;
 Best Local Similarity 63.6%; Pred. No. 63;
 Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 ttttgannnnnnnttgata 22
 |||||
 DB 2836 TGTGGAGCACTGGTTTGATA 2815

RESULT 8
 X34752
 ID X34752 standard; DNA; 396 BP.
 XX AC X34752;
 XX 05-JUL-1999 (first entry)
 XX DNA encoding OMP-1T protein.
 XX Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
 KW detection; dog; ss.
 XX Ehrlichia chaffeensis.
 XX W09913720-A1.
 XX 25-MAR-1999.
 XX 18-SEP-1998; 98WO-US19600.
 XX 19-SEP-1997; 97US-0059353.
 XX (OHIS) UNIV OHIO STATE.
 XX Ohashi N, Rikihisa Y;
 XX WPI; 1999-254290/21.
 XX P-PSDB; Y06952.
 XX Novel outer membrane proteins from Ehrlichia chaffeensis and
 XX Ehrlichia canis
 XX Disclosure; Fig 12A; 55pp; English.

XX The invention provides isolated outer membrane proteins (OMP) from
 CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
 CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
 CC in Y06943-958. The E. canis proteins form part of the P30 family and
 CC consist of proteins shown in Y06959-970. The proteins and genes are used
 CC to detect E. chaffeensis in patients and E. canis in dogs.
 XX Sequence 396 BP; 147 A; 49 C; 58 G; 142 T; 0 other;

Query Match 59.1%; Score 13; DB 20; Length 396;
 Best Local Similarity 61.9%; Pred. No. 1.7e+02;
 Matches 13; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 2 gttggannnnnnnttgata 22
 |||||
 DB 193 gttggaatgagttattgata 213

RESULT 9
 Z16735
 ID 216735 standard; cDNA; 828 BP.
 XX
 AC
 Z16735;
 XX
 DT 12-OCT-1999 (first entry)
 XX
 DE Human gene expression product cDNA sequence SEQ ID NO:4205.
 XX
 KW Human; gene; gene expression product; diagnosis; therapy; probe;
 KW detection; mapping; tissue typing; profiling; forensic; cancer;
 KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
 XX
 OS Homo sapiens.
 XX
 PN W0938972-A2.
 XX
 PD 05-AUG-1999.
 XX
 PF 28-JAN-1999; 99WO-US01619.
 XX
 PR 03-APR-1998; 98US-0080666.
 PR 28-JAN-1998; 98US-0072910.
 PR 24-FEB-1998; 98US-0075954.
 PR 31-MAR-1998; 98US-0080114.
 PR 03-APR-1998; 98US-0080515.
 XX
 PA (CHIR) CHIRON CORP.
 PA (HYSE-) HYSEQ INC.
 XX
 PI Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
 PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
 PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
 PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
 PI Stache-Crain B, Sudduth-Klinger J, Williams LT;
 XX
 DR WPI; 1999-494092/41.
 XX
 PT Novel human genes and their expression products which are
 PT differentially expressed in different cell types
 XX
 PS Claim 1; Page 1992; 2479pp; English.
 XX
 CC The present invention describes a library of human polynucleotides
 CC comprising the sequences given in 21532 to 21779. Also described is a
 CC method of detecting differentially expressed genes correlated with the
 CC cancerous state of a mammalian cell, comprising detecting at least one
 CC differentially expressed gene product in a test sample from a cell
 CC suspected of being cancerous, where the gene product is encoded by one
 CC of the 5248 polynucleotide sequences given in 21532 to 21779. The
 CC polynucleotides can be used as a source of primers and probes, which can
 CC be used for a variety of purpose, e.g. detection of expression levels,
 CC mapping, tissue typing or profiling, forensics, genetic analysis and
 CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
 CC can be used for raising antibodies for experimental, diagnostic and
 CC therapeutic purposes. The polynucleotides may also be used to construct
 CC arrays for diagnostics (which may be used to determine function of an
 CC encoded protein); and to detect differences in expression levels between
 CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
 CC identify a genetic predisposition or susceptibility to a disease such as
 CC cancer). The polynucleotides of the invention are especially used in the
 CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
 CC and lung cancer. The polynucleotides can also be used to screen for
 CC peptide analogues and antagonists.
 XX
 SQ Sequence 828 BP; 189 A; 155 C; 172 G; 227 T; 85 other;

Query Match 59.1%; Score 13; DB 20; Length 828;
 Best Local Similarity 59.1%; Pred. No. 1.8e+02;
 Matches 13; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 tgttgannnnnnnnntttgata 22
 |||||
 Db 199 tgttgagaattcatttgnta 220
 |||||
 RESULT 10
 X20676/c
 ID X20676 standard; DNA; 1079 BP.
 XX
 AC X20676;
 XX
 DT 05-MAY-1999 (first entry)
 XX
 DE Polynucleotide sequence from the genome of Treponema pallidum.
 XX
 KW Treponema pallidum infection; syphilis; Borrelia infection; animal;
 KW enzyme production; ds.
 XX
 OS Treponema pallidum.
 XX
 PN W09859034-A2.
 XX
 PD 30-DEC-1998.
 XX
 PF 23-JUN-1998; 98WO-US13041.
 XX
 PR 24-JUN-1997; 97US-0050667.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Fraser CM;
 XX
 DR WPI; 1999-081273/07.
 XX
 PT New isolated Treponema pallidum nucleic acids - used to develop
 PT products for the detection, diagnosis, characterisation, prevention
 PT and therapy of T. pallidum infections, particularly syphilis
 XX
 PS Claim 1; Page 842-843; 1150pp; English.
 PS
 CC X20500-21243 represent polynucleotide sequences from the genome of
 CC Treponema pallidum. The sequences can be used for detection,
 CC diagnosis, characterisation, prevention and therapy for T. pallidum
 CC infections, particularly syphilis. They can also be used for detecting
 CC diseases related to Borrelia infections in animals, and for the
 CC production of biosynthetic products such as enzymes.
 XX
 SQ Sequence 1079 BP; 397 A; 280 C; 181 G; 221 T; 0 other;

Query Match 59.1%; Score 13; DB 20; Length 1079;
 Best Local Similarity 61.9%; Pred. No. 1.9e+02;
 Matches 13; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 2 gttggannnnnnntttgata 22
 |||||
 Db 1022 GTTGGATTAAAGTGTTTGATA 1002
 |||||
 RESULT 11
 V18099
 ID V18099 standard; DNA; 1482 BP.
 XX
 AC V18099;
 XX
 DT 04-SEP-1998 (first entry)
 XX
 DE Nucleotide sequence of breakpoint region on chromosome 19.
 XX
 KW Hydronephrosis gene; HNG gene; USF2 gene; renal disease; renal aplasia;
 KW vesical-ureteral reflux; pelvi-ureteral junction obstruction;
 KW multicystic renal dysplasia; renal agenesis; hydronephrosis;

XX WO9901473-A2.
 XX 14-JAN-1999.
 XX PF 03-JUL-1998; 98WO-GB01974.
 XX PR 03-JUL-1997; 97US-0887534.
 XX (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.
 XX Holden DW;
 XX WPI; 1999-105999/09.
 XX P-PSDB; W97694.
 XX Inhibition of virulence genes from *Staphylococcus aureus* - useful
 PT for, e.g. screening for potential anti-microbial agents
 PT
 PS Claim 1; Page 86-89; 203pp; English.
 XX This is the nucleotide sequence of a virulence gene identified in
 CC *Staphylococcus aureus* mutant P15C31. *S. aureus* genes (see X07088-136)
 CC associated with virulence were identified by signature-tagged
 CC mutagenesis in which mutants containing a chromosomal insertion of
 CC a signature tagged transposon were generated, mutants with
 CC attenuated virulence were identified in a mouse model of bacteraemia,
 CC and the nucleotide sequences of the regions flanking the transposon
 CC insertion sites of these mutants were determined. Database sequence
 CC comparisons were performed to identify the virulence genes and to
 CC determine the possible function of their protein products (see also
 CC W97680-724). The P15C31 virulence gene product (see W97694) was
 CC identified as phosphoribosylformylglycinamide decarboxylase PurL,
 CC an enzyme involved in purine biosynthesis. A claimed method of
 CC identifying an antibacterial agent involves assaying potential agents
 CC for the ability to interfere with the expression of *S. aureus*
 CC virulence gene products. Also new is a *S. aureus* organism containing
 CC a functional mutation in one of the virulence genes, and its use in
 CC vaccine compositions.
 XX
 XX Sequence 2190 BP; 753 A; 309 C; 461 G; 666 T; 1 other;
 SQ

Query Match 59.1%; Score 13; DB 20; Length 2190;
 Best Local Similarity 61.9%; Pred. No. 2.1e+02;
 Matches 13; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 tgttggannnnnnntttgat 21
 ||||| |||||
 Db 1650 tgttgaatgtaggtttgat 1670

RESULT 14
 T67197
 ID T67197 standard; cDNA; 2508 BP.
 XX
 XX AC T67197;
 XX
 XX 03-SEP-1997 (first entry)
 XX
 XX Zebrafish retinoid X receptor RXR epsilon cDNA.
 DE
 DE Retinoid X receptor; RXR epsilon; zebrafish; ds.
 KW
 KW Brachydanio rerio.
 XX
 XX Key Location/Qualifiers
 FT CDS 388..1704
 FT /*tag= a
 XX
 XX CA2177642-A.
 XX
 XX 06-DEC-1996.
 PD

XX 29-MAY-1996; 96CA-2177642.
 XX
 XX 05-JUN-1995; 95US-0462182.
 XX
 XX (TOOH) UNIV QUEENS KINGSTON.
 XX
 XX Jones B, Ohno C, Petkovich M;
 PI
 XX WPI; 1997-298660/28.
 DR P-PSDB; W18033.
 XX
 XX Isolated retinoid X receptor protein - having amino acid insert in
 PT ligand binding domain, useful to identify specific target genes
 PT implicated in retinoid responses important in disease states
 XX
 XX Claim 8; Page 29-30; 42pp; English.
 PS
 XX cDNA clones (T67196 and T67197) respectively code for novel
 CC retinoid X receptors RXR delta (W18032) and RXR epsilon (W18033)
 CC that exhibit a high degree of amino acid conservation with other
 CC vertebrate RXRs but which represent unique subtypes defined by an
 CC additional 14-amino acid segment in their ligand binding domains.
 CC The RXR epsilon clone was isolated by screening adult and post-
 CC somitogenesis zebrafish cDNA libraries with a probe corresponding
 CC to the Drosophila melanogaster FTZ-F1alpha DNA binding domain.
 CC The isolated clones can be used to produce RXR delta and epsilon
 CC polypeptides useful for modifying retinoid activity and identifying
 CC specific target genes implicated in retinoid responses important in
 CC disease states.
 XX
 XX Sequence 2508 BP; 600 A; 652 C; 646 G; 610 T; 0 other;
 SQ

Query Match 59.1%; Score 13; DB 18; Length 2508;
 Best Local Similarity 61.9%; Pred. No. 2.1e+02;
 Matches 13; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 tgttggannnnnnntttgat 21
 ||||| |||||
 Db 2424 tgttgacatgagttttgat 2444

RESULT 15
 Z94941/c
 ID Z94941 standard; cDNA; 3519 BP.
 XX
 XX AC Z94941;
 XX
 XX 01-AUG-2000 (first entry)
 XX
 XX Human carbohydrate-associated protein CRBAP-1 cDNA.
 DE
 DE CRBAP-1; carbohydrate-associated protein 1; human;
 KW autoimmune disorder; inflammation; gastrointestinal disorder;
 KW infection; reproductive disorder; neurological disorder;
 KW eye disorder; cell proliferation; cancer; diagnosis; gene therapy;
 KW ss.
 XX
 XX Homo sapiens.
 OS
 XX Key Location/Qualifiers
 FT CDS 42..560
 FT /*tag= a
 FT variation replace(428,C)
 FT /*tag= b
 FT /*frequency= "0.10"
 FT /note= "a polynucleotide containing C at position
 FT 428 is specifically claimed in Claimed 21"
 XX
 XX WO200018922-A2.
 XX
 XX 06-APR-2000.
 PD

Search completed: March 27, 2001, 08:23:01
Job time: 5303 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 27, 2001, 08:19:22 ; Search time 132.3 Seconds
(without alignments)
26.799 Million cell updates/sec

Title: US-09-380-826A-6
Perfect score: 22
Sequence: 1 ttgtggannnnnnnttgata 22

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 280836 seqs, 80580151 residues

Total number of hits satisfying chosen parameters: 561672

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6_COMB.seq: *
4: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq: *
5: /cgn2_6/ptodata/2/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
C 1	14	63.6	3666	2	US-08-682-517-13 Sequence 13, Appl
C 2	14	63.6	3666	2	US-08-682-517-14 Sequence 14, Appl
C 3	14	63.6	4197	2	US-08-682-517-7 Sequence 7, Appl
C 4	14	63.6	4197	2	US-08-682-517-8 Sequence 8, Appl
C 5	13	59.1	5253	3	US-08-714-918-19 Sequence 19, Appl
C 6	12.4	56.4	1760	1	US-08-413-118-118 Sequence 118, App
C 7	12.4	56.4	1760	3	US-08-473-446-118 Sequence 118, App
C 8	12.4	56.4	1800	1	US-08-752-238-2 Sequence 2, Appl
C 9	12.4	56.4	1800	3	US-09-085-603B-2 Sequence 2, Appl
C 10	12.4	56.4	1800	3	US-09-031-897-6 Sequence 6, Appl
C 11	12.4	56.4	2730	1	US-08-339-129-1 Sequence 1, Appl
C 12	12.4	56.4	2761	1	US-08-752-238-1 Sequence 1, Appl
C 13	12.4	56.4	2761	3	US-09-085-603B-1 Sequence 1, Appl
C 14	12.4	56.4	2761	3	US-09-031-897-5 Sequence 5, Appl
C 15	12.4	56.4	3438	3	US-08-613-009A-1 Sequence 1, Appl
C 16	12	54.5	1130	1	US-07-864-004B-1 Sequence 1, Appl
C 17	12	54.5	1130	1	US-08-251-937A-1 Sequence 1, Appl
C 18	12	54.5	1130	1	US-08-212-133A-5 Sequence 5, Appl
C 19	12	54.5	1130	1	US-08-474-503-3 Sequence 3, Appl
C 20	12	54.5	1130	2	US-08-670-707A-3 Sequence 3, Appl
C 21	12	54.5	1130	4	PCT-US93-03275-1 Sequence 1, Appl
C 22	12	54.5	1130	4	PCT-US94-13200-3 Sequence 3, Appl
C 23	12	54.5	1623	1	US-08-121-202-3 Sequence 3, Appl
C 24	12	54.5	2026	3	US-08-755-587-26 Sequence 26, Appl
C 25	12	54.5	2625	1	US-08-468-036-2 Sequence 2, Appl
C 26	12	54.5	2625	2	US-08-376-843-2 Sequence 2, Appl
C 27	12	54.5	3284	3	US-09-136-652-1 Sequence 1, Appl
C 28	12	54.5	4334	2	US-08-670-707A-38 Sequence 38, Appl

C 29 12 54.5 4931 3 US-09-058-489-20 Sequence 20, Appl
C 30 12 54.5 6057 3 US-08-362-525-1 Sequence 1, Appl
C 31 12 54.5 6402 2 US-08-670-707A-36 Sequence 36, Appl
C 32 12 54.5 6476 3 US-09-058-489-21 Sequence 21, Appl
C 33 12 54.5 6727 3 US-08-629-643A-5 Sequence 5, Appl
C 34 12 54.5 6727 3 US-09-280-799-1 Sequence 1, Appl
C 35 12 54.5 13011 2 US-08-791-849A-14 Sequence 14, Appl
C 36 12 54.5 40352 3 US-08-846-111D-15 Sequence 15, Appl
C 37 11.4 51.8 21 2 US-08-639-501-23 Sequence 23, Appl
C 38 11.4 51.8 21 2 US-09-044-946-23 Sequence 23, Appl
C 39 11.4 51.8 21 3 US-09-044-908-23 Sequence 24, Appl
C 40 11.4 51.8 27 1 US-08-419-009-24 Sequence 24, Appl
C 41 11.4 51.8 293 3 US-08-866-340-13 Sequence 13, Appl
C 42 11.4 51.8 336 3 US-09-184-658-11 Sequence 11, Appl
C 43 11.4 51.8 485 1 US-08-419-009-19 Sequence 19, Appl
C 44 11.4 51.8 485 1 US-08-419-009-20 Sequence 20, Appl
C 45 11.4 51.8 485 1 US-08-419-009-21 Sequence 21, Appl

ALIGNMENTS

RESULT 1
US-08-682-517-13/c.
; Sequence 13, Application US/08682517
; Patent No. 5874267
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Expression of surface layer proteins
; NUMBER OF SEQUENCES: 25
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/682,517
; FILING DATE:
; CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3666 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-682-517-13

Query Match 63.6%; Score 14; DB 2; Length 3666;
Best Local Similarity 63.6%; Pred. No. 25;
Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 ttgtggannnnnnnttgata 22
|||||||
Db 2652 TGTGGAGCACTGGTTTGATA 2631

RESULT 2
US-08-682-517-14/c.
; Sequence 14, Application US/08682517
; Patent No. 5874267
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Expression of surface layer proteins
; NUMBER OF SEQUENCES: 25
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/682,517

```
;
; FILING DATE:
; CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3666 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3666
US-08-682-517-14
```

```
Query Match 63.6%; Score 14; DB 2; Length 3666;
Best Local Similarity 63.6%; Pred. No. 25;
Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
```

```
QY 1 ttttgannnnnnntttgata 22
|||||||
Db 2652 TGTGGAGCAACTGGTTTGATA 2631
```

RESULT 3

```
US-08-682-517-7/c
; Sequence 7, Application US/08682517
; Patent No. 5874267
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Expression of surface layer proteins
; NUMBER OF SEQUENCES: 25
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/682,517
; FILING DATE:
; CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4197 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-682-517-7
```

```
Query Match 63.6%; Score 14; DB 2; Length 4197;
Best Local Similarity 63.6%; Pred. No. 25;
Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
```

```
QY 1 ttttgannnnnnntttgata 22
|||||||
Db 2836 TGTGGAGCAACTGGTTTGATA 2815
```

RESULT 4

```
US-08-682-517-8/c
; Sequence 8, Application US/08682517
; Patent No. 5874267
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Expression of surface layer proteins
; NUMBER OF SEQUENCES: 25
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
```

```
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/682,517
; FILING DATE:
; CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4197 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Bacillus sphaericus
; INDIVIDUAL ISOLATE: P-1
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 95...3850
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 185...3850
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 95...184
US-08-682-517-8
```

```
Query Match 63.6%; Score 14; DB 2; Length 4197;
Best Local Similarity 63.6%; Pred. No. 25;
Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
```

```
QY 1 ttttgannnnnnntttgata 22
|||||||
Db 2836 TGTGGAGCAACTGGTTTGATA 2815
```

RESULT 5

```
US-08-714-918-19
; Sequence 19, Application US/08714918
; Patent No. 6037123
; GENERAL INFORMATION:
; APPLICANT: Benton, Bret
; APPLICANT: Lee, Ving
; APPLICANT: Malouin, Francois
; APPLICANT: Martin, Patrick K.
; APPLICANT: Schmid, Molly B.
; APPLICANT: Sun, Dongxu
; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS ANTIBACTERIAL
; TARGET GENES
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C.-DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/714,918
; FILING DATE: September 13, 1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/009,102
; FILING DATE: December 22, 1995
; APPLICATION NUMBER: 60/003,798
; FILING DATE: September 15, 1995
; ATTORNEY/AGENT INFORMATION:
```

```

; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 222/005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5253 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-714-918-19

Query Match 59.1%; Score 13; DB 3; Length 5253;
Best Local Similarity 61.9%; Pred. No. 94;
Matches 13; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 2 gttggannnnnnnttgata 22
|||||
Db 2895 GTTCGATGTAGTCTTTTGATA 2915

RESULT 6
US-08-413-118-118
; Sequence 118, Application US/08413118
; Patent No. 5688920
; GENERAL INFORMATION:
; APPLICANT: PAOLETTI, ENZO
; APPLICANT: LIMBACH, KEITH J.
; TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES OF
; CANINE HERPESVIRUS gB, gC, AND gD AND USES THEREFOR
; NUMBER OF SEQUENCES: 128
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CURTIS, MORRIS & SAFFORD, P.C.
; STREET: 530 FIFTH AVENUE, 25TH FLOOR
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/413,118
; FILING DATE: 29-MAR-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/220,151
; FILING DATE: 30-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: FROMMER, WILLIAM S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454310-2670
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; INFORMATION FOR SEQ ID NO: 118:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1760 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-413-118-118

Query Match 56.4%; Score 12.4; DB 1; Length 1760;
Best Local Similarity 59.1%; Pred. No. 1.7e+02;
Matches 13; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 tgttgannnnnnnttgata 22
|||||
Db 1076 TGTTGGATAACTATTTTATA 1097

RESULT 7
US-08-473-446-118
; Sequence 118, Application US/08473446
; Patent No. 6017542
; GENERAL INFORMATION:
; APPLICANT: PAOLETTI, ENZO
; APPLICANT: LIMBACH, KEITH J.
; TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES OF
; CANINE HERPESVIRUS gB, gC, AND gD AND USES THEREFOR
; NUMBER OF SEQUENCES: 128
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CURTIS, MORRIS & SAFFORD, P.C.
; STREET: 530 FIFTH AVENUE, 25TH FLOOR
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/473,446
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/413,118
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: FROMMER, WILLIAM S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454310-2670
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; INFORMATION FOR SEQ ID NO: 118:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1760 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-473-446-118

Query Match 56.4%; Score 12.4; DB 3; Length 1760;
Best Local Similarity 59.1%; Pred. No. 1.7e+02;
Matches 13; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 tgttgannnnnnnttgata 22
|||||
Db 1076 TGTTGGATAACTATTTTATA 1097

RESULT 8
US-08-752-238-2/c
; Sequence 2, Application US/08752238
; Patent No. 5804418
; GENERAL INFORMATION:
; APPLICANT: Lambowitz Dr., Alan M
; APPLICANT: Mohr Dr., Georg
; APPLICANT: Saidanha Dr., Roland
; APPLICANT: Matsura Dr., Manabu
; TITLE OF INVENTION: Method for Preparing Nucleotide
; TITLE OF INVENTION: Integrase
```

NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: CALFEE, HALTER & GRISWOLD
STREET: 800 Superior Avenue
CITY: Cleveland
STATE: Ohio
COUNTRY: US
ZIP: 44114
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/752,238
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Golrick, Mary E.
REGISTRATION NUMBER: 34,829
REFERENCE/DOCKET NUMBER: 24671/00103
TELECOMMUNICATION INFORMATION:
TELEPHONE: (216) 622-8458
TELEFAX: (216) 241-0816
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1800 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1800
US-08-752-238-2

Query Match 56.4%; Score 12.4; DB 1; Length 1800;
Best Local Similarity 59.1%; Pred. No. 1.7e+02;
Matches 13; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 ttttggannnnnnnttgata 22
| | | | |
Db 139 TATTGGAATATAAATTTTGATA 118

RESULT 9
US-09-085-603B-2/c
Sequence 2, Application US/09085603B
Patent No. 6001608
GENERAL INFORMATION:
APPLICANT: Lambowitz Dr., Alan M
APPLICANT: Mohr Dr., Georg
APPLICANT: Saldanha Dr., Roland
APPLICANT: Matsuura Dr., Manabu
APPLICANT: Yang Dr., Jiam
APPLICANT: Zimmerly Dr., Steven
APPLICANT: Guo Dr., Huatao
APPLICANT: Beall Dr., Clifford J.
TITLE OF INVENTION: Methods of Making an Rnp Particle
TITLE OF INVENTION: Having Nucleotide Integrase Activity
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: CALFEE, HALTER & GRISWOLD LLP
STREET: 800 Superior Avenue
CITY: Cleveland
STATE: Ohio
COUNTRY: US
ZIP: 44114
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/085,603B
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Docherty, Pamela A.
REGISTRATION NUMBER: 40,591
REFERENCE/DOCKET NUMBER: 24671/04000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (216) 622-8416
TELEFAX: (216) 241-0816
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1800 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1800
US-09-085-603B-2
Query Match 56.4%; Score 12.4; DB 3; Length 1800;
Best Local Similarity 59.1%; Pred. No. 1.7e+02;
Matches 13; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 ttttggannnnnnnttgata 22
| | | | |
Db 139 TATTGGAATATAAATTTTGATA 118

RESULT 10
US-09-031-897-6/c
Sequence 6, Application US/09031897
Patent No. 6027895
GENERAL INFORMATION:
APPLICANT: Lambowitz, Alan
APPLICANT: Mohr, Georg
APPLICANT: Zimmerly, Steven
APPLICANT: Guo, Huatao
TITLE OF INVENTION: Methods Cleaving DNA with Nucleotide
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Calfee, Halter & Griswold
STREET: 800 Superior Avenue, Suite 1400
CITY: Cleveland
STATE: Ohio
COUNTRY: US
ZIP: 44114
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/031,897
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Docherty, Pamela A.
REGISTRATION NUMBER: 40,591
REFERENCE/DOCKET NUMBER: 24671/00105
TELECOMMUNICATION INFORMATION:
TELEPHONE: (216) 622-8416
TELEFAX: (216) 241 0816
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1800 base pairs
TYPE: nucleic acid

```

: Patent No. 5804418
:
: GENERAL INFORMATION:
: APPLICANT: Lambowitz Dr., Alan M
: APPLICANT: Mohr Dr., Georg
: APPLICANT: Saldanha Dr., Roland
: APPLICANT: Matsuura Dr., Manabu
: TITLE OF INVENTION: Method for Preparing Nucleotide
: TITLE OF INVENTION: Integrase
: NUMBER OF SEQUENCES: 12
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: CALFEE, HALTER & GRISWOLD
: STREET: 800 Superior Avenue
: CITY: Cleveland
: STATE: Ohio
: COUNTRY: US
: ZIP: 44114
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/752,238
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Golrick, Mary E.
: REGISTRATION NUMBER: 34,829
: REFERENCE/DOCKET NUMBER: 24671/00103
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (216) 622-8458
: TELEFAX: (216) 241-0816
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2761 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
:
: US-08-752-238-1
:
: Query Match 56.4%; Score 12.4; DB 1; Length 28
: Best Local Similarity 59.1%; Pred. No. 1.8e+02;
: Matches 13; Conservative 0; Mismatches 9; Indels
:
: QY 1 tgttgagannnnnttgata-22
:
: Db 890 TATTGGAATATAAATTTGATA 869
:
: RESULT 13
: US-09-085-603B-1/G
: Sequence 1, Application US/09085603B
: Patent No. 6001608
:
: GENERAL INFORMATION:
: APPLICANT: Lambowitz Dr., Alan M
: APPLICANT: Mohr Dr., Georg
: APPLICANT: Saldanha Dr., Roland
: APPLICANT: Matsuura Dr., Manabu
: APPLICANT: Yang Dr., Jiam
: APPLICANT: Zimmerly Dr., Steven
: APPLICANT: Guo Dr., Huatao
: APPLICANT: Beall Dr., Clifford J.
: TITLE OF INVENTION: Methods of Making an Rnp Particle
: TITLE OF INVENTION: Having Nucleotide Integrase Activity
: NUMBER OF SEQUENCES: 23
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: CALFEE, HALTER & GRISWOLD LLP
: STREET: 800 Superior Avenue
: CITY: Cleveland
: STATE: Ohio
: COUNTRY: US

```

Qy 1 tggtggannnnnnnnnttgata 22
|||||
Db 202 TGTTGGCTTTTGCACCTTTGATA 181

Search completed: March 27, 2001, 08:19:23
Job time: 5150 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 27, 2001, 07:38:23 ; Search time 2517.78 Seconds
(without alignments)
61.230 Million cell updates/sec

Title: US-09-380-826A-6
Perfect score: 22
Sequence: 1 tgttgannnnnnntttgata 22

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues

Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: em_estba:*
42: em_estfun:*
43: em_esthum1:*

44: em_esthum2:*
45: em_esthum3:*
46: em_esthum4:*
47: em_esthum5:*
48: em_esthum6:*
49: em_esthum7:*
50: em_esthum8:*
51: em_esthum9:*
52: em_esthum10:*
53: em_esthum11:*
54: em_esthum12:*
55: em_esthum13:*
56: em_esthum14:*
57: em_esthum15:*
58: em_esthum16:*
59: em_esthum17:*
60: em_esthum18:*
61: em_esthum19:*
62: em_esthum20:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estov1:*
68: em_estov2:*
69: em_estpl1:*
70: em_estpl2:*
71: em_estpl3:*
72: em_estpl4:*
73: em_estpl5:*
74: em_estrol:*
75: em_estro2:*
76: em_estro3:*
77: em_estro4:*
78: em_estro5:*
79: em_estro6:*
80: em_estro7:*
81: em_estro8:*
82: em_estro9:*
83: em_estrol1:*
84: em_estrol2:*
85: em_estrol3:*
86: em_estrol4:*
87: gb_est41:*
88: gb_est42:*
89: gb_est43:*
90: gb_est44:*
91: gb_est45:*
92: gb_est46:*
93: gb_est47:*
94: gb_est48:*
95: gb_est49:*
96: gb_est50:*
97: gb_est51:*
98: gb_est52:*
99: gb_est53:*
100: gb_est54:*
101: gb_est55:*
102: gb_est56:*
103: gb_est57:*
104: gb_est67:*
105: gb_est68:*
106: gb_est69:*
107: gb_est70:*
108: gb_est71:*
109: gb_est72:*
110: gb_est73:*
111: gb_est74:*
112: em_esthum21:*
113: em_esthum22:*
114: em_esthum23:*
115: em_estom1:*
116: em_estom2:*

117: em_estpl6:*
 118: em_estpl7:*
 119: em_estpl8:*
 120: em_estpl9:*
 121: em_estpl10:*
 122: em_estpl11:*
 123: em_estpl12:*
 124: em_estpl13:*
 125: em_estpl14:*
 126: em_estpl15:*
 127: em_estpl16:*
 128: em_estpl17:*
 129: em_estpl18:*
 130: em_estpl19:*
 131: em_estpl20:*
 132: em_estpl21:*
 133: em_estpl22:*
 134: em_estpl23:*
 135: em_estpl24:*
 136: em_estpl25:*
 137: em_estpl26:*
 138: em_estpl27:*
 139: em_estpl28:*
 140: em_estpl29:*
 141: em_estpl30:*
 142: em_estpl31:*
 143: em_estpl32:*
 144: em_estpl33:*
 145: em_estpl34:*
 146: em_estpl35:*
 147: em_estpl36:*
 148: em_estpl37:*
 149: em_estpl38:*
 150: em_estpl39:*
 151: em_estpl40:*
 152: em_estpl41:*
 153: em_estpl42:*
 154: em_estpl43:*
 155: em_estpl44:*
 156: em_estpl45:*
 157: em_estpl46:*
 158: em_estpl47:*
 159: em_estpl48:*
 160: em_estpl49:*
 161: em_estpl50:*
 162: em_estpl51:*
 163: em_estpl52:*
 164: em_estpl53:*
 165: em_estpl54:*
 166: em_estpl55:*
 167: em_estpl56:*
 168: em_estpl57:*
 169: em_estpl58:*
 170: em_estpl59:*
 171: em_estpl60:*
 172: em_estpl61:*
 173: em_estpl62:*
 174: em_estpl63:*
 175: em_estpl64:*
 176: em_estpl65:*
 177: em_estpl66:*
 178: em_estpl67:*
 179: em_estpl68:*
 180: em_estpl69:*
 181: em_estpl70:*
 182: em_estpl71:*
 183: em_estpl72:*
 184: em_estpl73:*
 185: em_estpl74:*
 186: em_estpl75:*
 187: em_estpl76:*
 188: em_estpl77:*
 189: em_estpl78:*

190: gb_gss25:*
 191: gb_gss26:*
 192: gb_gss27:*
 193: gb_gss28:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
C 1	14	63.6	184	128	BB357781	BB357781 BB357781
C 2	14	63.6	210	134	BE059640	BE059640 sn35a01.y
C 3	14	63.6	228	138	BE956026	BE956026 UI-W-BH4-
C 4	14	63.6	235	102	BB170254	BB170254 BB170254
C 5	14	63.6	250	16	AI122094	AI122094 uc46f10.r
C 6	14	63.6	261	96	AW879900	AW879900 QV3-OT002
C 7	14	63.6	265	126	BB287901	BB287901 BB287901
C 8	14	63.6	298	8	AA515145	AA515145 ng68c01.s
C 9	14	63.6	332	105	BE202756	BE202756 EST402778
C 10	14	63.6	336	166	AZ037213	AZ037213 RPCI-23-3
C 11	14	63.6	359	37	AV546877	AV546877 AV546877
C 12	14	63.6	364	87	AW226052	AW226052 ST76608.P
C 13	14	63.6	371	151	AQ311167	AQ311167 CITBI-EI-
C 14	14	63.6	376	149	AQ102282	AQ102282 HS_3029.A
C 15	14	63.6	376	163	AQ444476	AQ444476 an37g11.J
C 16	14	63.6	378	88	AW240486	AW240486 ug35b10.x
C 17	14	63.6	389	8	AA488558	AA488558 ab37h10.r
C 18	14	63.6	399	144	R89220	R89220 yp99a12.t1
C 19	14	63.6	410	93	AW624588	AW624588 EST322533
C 20	14	63.6	410	149	AQ150667	AQ150667 HS_3197_B
C 21	14	63.6	416	9	AA555292	AA555292 nk82g03.s
C 22	14	63.6	427	159	AQ549780	AQ549780 RPCI-11-3
C 23	14	63.6	427	23	AI662442	AI662442 mt25a11.x
C 24	14	63.6	447	169	AZ273408	AZ273408 RPCI-23-1
C 25	14	63.6	459	136	BE854379	BE854379 ux29e12.y
C 26	14	63.6	467	28	AL370583	AL370583 MCB38F11
C 27	14	63.6	471	22	AI591928	AI591928 mt31b03.y
C 28	14	63.6	476	164	AQ914606	AQ914606 nbe00049C
C 29	14	63.6	481	146	T75954	T75954 10732.Lambd
C 30	14	63.6	482	174	B29252	B29252 T29J13TF.TA
C 31	14	63.6	489	105	BE196823	BE196823 ug71a09.y
C 32	14	63.6	490	108	BE448206	BE448206 ut60g10.y
C 33	14	63.6	493	10	AA637362	AA637362 vu09f03.r
C 34	14	63.6	496	159	AQ589820	AQ589820 HS_2136_B
C 35	14	63.6	497	107	BE353412	BE353412 EST353789
C 36	14	63.6	498	138	BF003327	BF003327 EST431825
C 37	14	63.6	521	93	AW649760	AW649760 EST328214
C 38	14	63.6	531	90	AW413577	AW413577 uc37a12.x
C 39	14	63.6	531	93	AW648609	AW648609 EST327159
C 40	14	63.6	531	160	AQ669223	AQ669223 HS_5415_B
C 41	14	63.6	536	97	AW945076	AW945076 EST337127
C 42	14	63.6	539	27	AI997301	AI997301 701553168
C 43	14	63.6	540	90	AW398351	AW398351 EST298198
C 44	14	63.6	544	38	AV678686	AV678686 AV678686
C 45	14	63.6	544	169	AZ267308	AZ267308 RPCI-23-4

ALIGNMENTS

RESULT 1
 BB357781/C

LOCUS

DEFINITION

musculus CDNA clone C030024G02 3', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

184 bp mRNA

BB357781

RIKEN full-length enriched,

adult male corpus striatum Mus

EST

12-JUL-2000

house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 184)

AUTHORS P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shigenoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomimaga, N., Toya, T., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yamane, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

TITLE RIKEN Mouse ESTs (Konno, H., et al.)

JOURNAL Unpublished (2000)

COMMENT Contact: Yoshihide Hayashizaki
Genome Exploration Research Group, Life Science Tsukuba Center,
Genome Science Laboratory
The Institute of Physical and Chemical Research (RIKEN), Genomic
Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: +81-298-36-9013
Fax: +81-298-36-9098
Email: genome-res@rtc.riken.go.jp/
URL: <http://genome.rtc.riken.go.jp/>
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Thermotabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh, M., Kitsunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (3), 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

FEATURES source

Location/Qualifiers

1..184
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="C030024G02"
/clone_lib="RIKEN full-length enriched, adult male corpus striatum"
/sex="male"
/tissue_type="corpus striatum"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGATTCGAGTTAAATTAATCCCTCCCCCCCC sequence [5' GAGAGAGATTCGAGTTAAATTAATTAATCCCTCCCCCCCC sequence [5' GAGAGAGATTCGAGTTAAATTAATTAATCCCTCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI"

BASE COUNT 51 a 34 c 39 g 60 t

ORIGIN

Query Match 63.6%; Score 14; DB 128; Length 184;
Best Local Similarity 63.6%; Pred No. 6.5e+02;
Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 tgttgannnnnnnnnttgata 22
|||||

Db 82 TGTGGAAAATGGCTTTTGATA 61
|||||

RESULT 2

LOCUS BE059640

DEFINITION BE059640.1 GI:8404006

ACCESSION BE059640

VERSION BE059640.1

KEYWORDS EST.

SOURCE soybean.

ORGANISM Glycine max

REFERENCE Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.

AUTHORS Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Corvelli, V., Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, B., Theising, B., Allen, M., Bowers, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.

TITLE Public Soybean EST Project

JOURNAL Unpublished (1999)

COMMENT Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com
Seq primer: -40RP from Gibco
High quality sequence stop: 165.
Location/Qualifiers

1..210
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl016-13441"
/clone_lib="Gm-cl016"
/tissue_type="Immature flowers of field grown plants"
/lab_host="XL10-Gold"
/note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2: XhoI; This cDNA library was constructed from mRNA isolated from immature flowers of field grown plants. The cDNA library was prepared using the Stratagene pBluescript II XR library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly (dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into XL10-Gold host cells. This library was constructed by Dr. Randy Shoemaker and Dr. John Erpelting."

BASE COUNT 57 a 38 c 55 g 60 t

ORIGIN

Query Match 63.6%; Score 14; DB 134; Length 210;
 Best Local Similarity 63.6%; Pred. No. 6.7e+02;
 Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 tgttgannnnnnnnnnttgata 22
 |||||
 Db 43 TGTGGATGTCACACGTTTGATA 64

RESULT

BE956026/c
 LOCUS BE956026 228 bp mRNA EST 04-OCT-2000
 DEFINITION UI-M-BH4-bav-a-11-0-UI.s1 NIH_BMAP_M_S5 Mus musculus cDNA clone

ACCESSION BE956026
 VERSION BE956026.1 GI:10600157
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 228)

AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene
 discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 COMMENT Contact: Chin, H
 National Institute of Mental Health
 6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
 20892-9643 USA
 Tel: 301 443 1706
 Fax: 301 443 9890

Email: mestr@mail.nih.gov
 The sequence contained an oligo-dT track that was present in the
 oligonucleotide that was used to prime the synthesis of first
 strand cDNA and therefore this may represent a bonafide poly A
 tail. The sequence tag present in the cDNA between the NotI site
 and the oligo-dT track served to identify it as a clone from the
 olfactory bulbs tissue cDNA Library Preparation: M.B. Soares Lab
 Clone distribution: Researchers may obtain BMAP cDNA clones from
 RESEARCH GENETICS. It should be noted that Bento Soares is
 generating a small number of additional specialized non-redundant
 arrays of BMAP cDNAs whose availability will be considered under
 appropriate and limited collaborative arrangements
 Seq primer: M13 Forward
 POLYA=yes.

FEATURES

source

Location/Qualifiers
 1..228
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UI-M-BH4-bav-a-11-0-UI"
 /clone_lib="NIH_BMAP_M_S5"
 /dev_stage="27-32 days"
 /lab_host="DH10B (Life Technologies)"
 /note="vector: pT73D-Pac (Pharmacia) with a modified
 polylinker: Site_1: Not I; Site_2: Eco RI; The
 NIH_BMAP_M_S5 library is a subtracted library of a series,
 ultimately derived from a mixture of individually tagged
 normalized libraries from ten regions of the mouse brain
 (cerebellum, brain stems, olfactory bulbs, hypothalamus,
 cortex, amygdala, basal ganglia, pineal gland, striatum,
 hippocampus) after a series of subtractions to reduce the
 representation of cDNAs from which ESTs had already been
 generated. For a detailed description of the library from
 which this clone was derived, please visit our web site
 at brainest.eng.uiowa.edu.
 TAG_LIB=NIH_BMAP_M_S5
 TAG_TISSUE=olfactory-bulbs

TAG_SEQ=CAAGG"

BASE COUNT 64 a 47 c 42 g 75 t
 ORIGIN

Query Match 63.6%; Score 14; DB 138; Length 228;
 Best Local Similarity 63.6%; Pred. No. 6.8e+02;
 Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 tgttgannnnnnnnnnttgata 22
 |||||
 Db 144 TGTGGAATCTGCTTTTGATA 123

RESULT

4
 BB170254/c

LOCUS BB170254

DEFINITION RIKEN full-length enriched, adult male hypothalamus Mus

musculus cDNA clone A230017D13 3', mRNA sequence.

ACCESSION BB170254

VERSION BB170254.1

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 235)

AUTHORS Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci

,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,

Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,

Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,

Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.

, Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.

, Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata

,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sobabe,Y., Sugahara,Y.

, Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tomimaga,N., Toya

,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.

, Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino

,M., Muramatsu,M. and Hayashizaki,Y.

RIKEN Mouse ESTs (Konno,H., et al.)

UNPUBLISHED (2000)

CONTACT: Yoshihide Hayashizaki

Genome Exploration Research Group, Life Science Tsukuba Center,

The Institute of Physical and Chemical Research (RIKEN), Genomic

Sciences Center

3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan

Tel: +81-298-36-9013

Fax: +81-298-36-9038

Email: genome-res@rtr.riken.go.jp,

URL: http://genome.rtr.riken.go.jp/

Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki

,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

Thermotabilization and thermoactivation of thermolabile enzymes by

trehalose and its application for the synthesis of full length

cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)

Itoh,M., Katsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,

Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki

,Y. and Hayashizaki,Y.

Automated filtration-based high-throughput plasmid preparation

system. Genome Res. 9 (5), 463-470 (1999)

Carninci,P. and Hayashizaki,Y.

High-efficiency full-length cDNA cloning. Methods Enzymol. 303,

19-44 (1999)

Please visit our web site (<http://genome.rtr.riken.go.jp>) for

further details.

Location/Qualifiers

1..235

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone="A230017D13"

/clone_lib="RIKEN full-length enriched, adult male

hypothalamus"

FEATURES

source

```

/sex="male"
/tissue_type="hypothalamus"
/lab_host="adult"
/notes="Site 1: Sali; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAAGGATCCCAAGAGCTCTTTTCTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATTCTCGATTAAATTAATCCGCCCCGCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."
BASE COUNT      84 a  47 c  31 g  73 t
ORIGIN

```

```

Query Match      63.6%; Score 14; DB 102; Length 235;
Best Local Similarity 63.6%; Pred. No. 6.9e+02;
Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 tgttgagannnnnnnnttgata 22
|||||
Db 66 TGTGGACTTAGTCTTGTGATA 45

```

```

RESULT 5
LOCUS      A1122094
DEFINITION uc46f10_r1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone IMAGE:1401067 5' similar to SW:1335_MTCGE P47577 HYPOTHETICAL GTP-BINDING PROTEIN MG335. ;, mRNA sequence.
ACCESSION  A1122094
VERSION     A1122094.1 GI:3522418
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 250)
Marra M., Hillier L., Allen M., Bowles M., Dietrich N., Dubuque T., Geisel S., Kucaba T., Lacy M., Le M., Martin J., Morris M., Schellenberg K., Steptoe M., Tan F., Underwood K., Moore B., Theising B., Wylie T., Lennon G., Soares B., Wilson R. and Waterston R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu

```

```

This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MG1:912783
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 225.
Location/Qualifiers
1. .250
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:1401067"
/clone_lib="Soares_mammary_gland_NMLMG"

```

FEATURES

source

```

/note="Organ: ovary; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 /716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

```

```

/sex="female (lactating)"
/tissue_type="mammary gland"
/lab_host="PH10B"
/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from mammary gland tissue from a lactating female, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      48 a  79 c  75 g  48 t
ORIGIN

```

```

Query Match      63.6%; Score 14; DB 16; Length 250;
Best Local Similarity 63.6%; Pred. No. 6.9e+02;
Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 tgttgagannnnnnnnttgata 22
|||||
Db 211 TGTGGAAAATCTCTTGTGATA 232

```

```

RESULT 6
LOCUS      AW879900/c
DEFINITION QV3-OT0029-290300-134-b02 OT0029 Homo sapiens cDNA, mRNA sequence.
ACCESSION  AW879900
VERSION     AW879900.1 GI:8041910
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 261)
Dias Neto E., Garcia Correa R., Verjovski-Almeida S., Briones M.R., Nagai M.A., da Silva W. Jr., Zago M.A., Bordin S., Costa F., Goldman G.H., Carvalho A.F., Matsukuma A., Baia G.S., Simpson D.H., Brunstein A., deOliveira P.S., Bucher P., Jorgeneel C.V., O'Hare M.J., Soares F., Brentani R.R., Reis L.F., de Souza S.J. and Simpson A.J.
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br

```

```

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl1=6t2-QV3-OT0029-290300-134-b02&t3=2000-03-29&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 261.
Location/Qualifiers
1. .261
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="OT0029"
/dev_stage="Adult"
/note="Organ: ovary; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 /716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

```

```

TITLE
JOURNAL
MEDLINE
COMMENT

```

FEATURES

source

•
J.


```

BASE COUNT      76 a    63 c    91 g    68 t
ORIGIN

Query Match      63.6%; Score 14; DB 8; Length 298;
Best Local Similarity 63.6%; Pred. No. 7.2e+02;
Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 tgttgannnnnnnttgata 22
|||||
Db 5 TGTGGAGATGACATTTGATA 26

RESULT 9
BE202756
LOCUS      BE202756      332 bp      mRNA      EST      07-SEP-2000
DEFINITION EST402778 KV1 Medicago truncatula cDNA clone pkV1-3023, mRNA
sequence.
ACCESSION  BE202756
VERSION    BE202756
KEYWORDS   BE202756.1 GI:8746021
SOURCE     barrel medic.
ORGANISM   Medicago truncatula
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
            Fabales; Fabaceae; Papilionoideae; Medicago.
REFERENCE  1 (bases 1 to 332)
AUTHORS   Vandenbosch,K., Endre,G., Hur,J., Moore,J., Beremand,P., Ellis,L.,
            Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and
            Fraser,C.M.
TITLE     ESTs from roots of Medicago truncatula 24 hours after inoculation
            with Sinorhizobium meliloti
JOURNAL   Unpublished (1999)
COMMENT   Contact: Vandenbosch K
            Department of Biology
            Texas A&M University
            College Station, TX 77843-3258, USA
            Tel: 409 845 7707
            Fax: 409 845 2891
            Email: kate@mail.bio.tamu.edu
            Texas A&M University:T262522e
            TIGR sequence name:MTIAE967K
            More information is available at:
            http://chrysie.tamu.edu/medicago
            Seq primer: SKnod (CTA gAA CTA gTg gAT CC).
FEATURES
            Location/Qualifiers
            1..332
                /organism="Medicago truncatula"
                /cultivar="genotype A17"
                /db_xref="taxon:3880"
                /clone="pkV1-3023"
                /clone_lib="KV1"
                /tissue_type="Seedling roots"
                /dev_stage="24 hours post-inoculation with Sinorhizobium
                meliloti"
                /lab_host="E. coli strain XL0LR"
                /note="Vector: pBluescript SK -; Site_1: EcoRI; Site_2:
                XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
                was directionally ligated into the Unizap XR vector from
                Stratagene and packaged using Gigapack III Gold packaging
                extracts. Plasmids containing cDNA inserts were excised
                from the recombinant lambda-zap phage using Ex-assist
                helper phage and propagated in XL0LR cells."
BASE COUNT      95 a    50 c    96 g    91 t
ORIGIN

Query Match      63.6%; Score 14; DB 105; Length 332;
Best Local Similarity 63.6%; Pred. No. 7.4e+02;
Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 tgttgannnnnnnttgata 22
|||||

```

```

Db 157 TGTGTGATTGATGGGTTTGATA 178

RESULT 10
AZ037213
LOCUS      AZ037213      336 bp      DNA      GSS      01-MAR-2000
DEFINITION RPCI-23-364L15.TV RPCI-23 Mus musculus genomic clone RPCI-23-364L15
            , DNA sequence.
ACCESSION  AZ037213
VERSION    AZ037213.1 GI:7125453
KEYWORDS   GSS.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 336)
AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
            ,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
            and Fraser,C.M.
TITLE     Mouse BAC End Sequences from Library RPCI-23
JOURNAL   Unpublished (1999)
COMMENT   Other_GSSs: RPCI-23-364L15.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 364 row: L column: 15
            Seq primer: T7
            Class: BAC ends.
FEATURES
            Location/Qualifiers
            1..336
                /organism="Mus musculus"
                /strain="C57BL/6J"
                /db_xref="taxon:10090"
                /clone="RPCI-23-364L15"
                /clone_lib="RPCI-23"
                /sex="Female"
                /lab_host="DH10B"
                /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                brain genomic DNA was isolated and partially digested
                with a combination of EcoRI and EcoRI Methylase. Size
                selected DNA was cloned into the pBACe3.6 vector at the
                EcoRI sites. The ligation products were transformed into
                DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT      97 a    37 c    45 g    157 t
ORIGIN

Query Match      63.6%; Score 14; DB 166; Length 336;
Best Local Similarity 63.6%; Pred. No. 7.4e+02;
Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 tgttgannnnnnnttgata 22
|||||
Db 268 TGTGGAATTTTCCTTTTGATA 289

RESULT 11
AV546877
LOCUS      AV546877      359 bp      mRNA      EST      07-SEP-2000
DEFINITION AV546877 Arabidopsis thaliana roots Columbia Arabidopsis thaliana
            cDNA clone R2L2la12F 3', mRNA sequence.
ACCESSION  AV546877

```

```

VERSION AV546877.1 GI:8718291
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
REFERENCE Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
AUTHORS Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
TITLE Brassicales; Brassicaceae; Arabidopsis.
JOURNAL Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.
MEDLINE A large scale analysis of cDNA in Arabidopsis thaliana: Generation
COMMENT of 12,028 non-redundant expressed sequence tags from normalized and
size-selected cDNA libraries
DNA Res. 7, 175-180 (2000)
20363093
Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
FEATURES
    source
        Location/Qualifiers
            1..359
                /organism="Arabidopsis thaliana"
                /strain="Columbia"
                /db_xref="taxon:3702"
                /clone_lib="RZL21a12F"
                /clone_lib="Arabidopsis thaliana roots Columbia"
                /tissue_type="roots"
                /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                XhoI"
BASE COUNT 107 a 65 c 47 g 140 t
ORIGIN

Query Match 63.6%; Score 14; DB 37; Length 359;
Best Local Similarity 63.6%; Pred. No. 7.5e+02;
Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 tgttgannnnnnntttgata 22
|||||
Db 8 TGTGGAAGTATTATTGATA 29

RESULT 12
AW226052
LOCUS 364 bp mRNA EST 10-DEC-1999
DEFINITION ST76G08 Pine Triplex shoot tip library Pinus taeda cDNA clone
ST76G08, mRNA sequence.
ACCESSION AW226052
VERSION AW226052.1 GI:6555348
KEYWORDS EST.
SOURCE loblolly pine.
ORGANISM Pinus taeda
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
REFERENCE 1 (bases 1 to 364)
AUTHORS Whetten, R.W., Kinlaw, C.S., Retzel, E. and Sederoff, R.R.
TITLE The Pine Gene Discovery Project
JOURNAL Unpublished (1999)
COMMENT Contact: Ross Whetten
Forest Biotechnology Group
North Carolina State University
Dept. of Forestry, NC State University, 6113 Jordan Hall, Raleigh
NC, 27695-8008
Tel: 919-515-7800
Fax: 919-515-7801
Email: rosswhet@unity.ncsu.edu
Seq primer: 5' lambda Triplex2 Sequencing Primer.
Location/Qualifiers
    1..364
        /organism="Pinus taeda"
        /db_xref="taxon:3352"
        /clone="ST76G08"
        /clone_lib="Pine Triplex shoot tip library"

```

```

/lab_host="E. coli BM25.8"
/note="Organ: shoot tips; Vector: Lambda Triplex; Site_1:
SfiI (A); Site_2: SfiI (B); Shoot tips (approx. 2 cm from
apex) were collected during the spring, frozen and used
for mRNA isolation. The SMART-PCR method (Clontech) was
used to prepare a library from 1 ug total RNA, using the
Lambda triplex vector. Plasmid subclones in pTriplex were
recovered by cre-lox excision in E. coli strain BM25.8 and
sequenced from the 5' end."
BASE COUNT 106 a 48 c 74 g 136 t
ORIGIN

Query Match 63.6%; Score 14; DB 87; Length 364;
Best Local Similarity 63.6%; Pred. No. 7.5e+02;
Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 tgttgannnnnnntttgata 22
|||||
Db 99 TGTGATGAAGTTTTCGATA 120

RESULT 13
AQ311167/c
LOCUS 371 bp DNA GSS 22-DEC-1998
DEFINITION CITBI-E1-2525E11.TR CITBI-E1 Homo sapiens genomic clone 2525E11,
DNA sequence.
ACCESSION AQ311167
VERSION AQ311167.1 GI:4042980
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
REFERENCE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS 1 (bases 1 to 371)
Adams, M.D., Rounsley, S.D., Zhao, S., Bass, S., Linher, K., Golden, K.,
Berry, K., Granger, D., Suh, E., Wible, C., Shizuya, H., Simon, M. and
Venter, J.C.
TITLE Use of a random human BAC End Sequence Database for Sequence-Ready
JOURNAL Map Building
COMMENT Unpublished (1998)
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13 Reverse
Class: BAC ends.
FEATURES
    source
        Location/Qualifiers
            1..371
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="2525E11"
                /clone_lib="CITBI-E1"
                /sex="male"
                /cell_type="sperm"
                /note="Vector: pBelBAC11; Site_1: EcoRI; Site_2: EcoRI;
                Caltech Human BAC Library D"
BASE COUNT 131 a 69 c 85 g 86 t
ORIGIN

Query Match 63.6%; Score 14; DB 151; Length 371;
Best Local Similarity 63.6%; Pred. No. 7.5e+02;
Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 tgttgannnnnnntttgata 22
|||||

```

Db 276 TGTTGGAATTACAGTTTGATA 255

RESULT 14
AQ102282/c

LOCUS
DEFINITION

AQ102282 376 bp DNA GSS 27-AUG-1998
HS_3029_A2_G04_MF CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate=3029 Col=8 Row=M, DNA sequence.

ACCESSION
AQ102282

VERSION
AQ102282.1 GI:3473311

KEYWORDS
GSS.

SOURCE
human.

ORGANISM
Homo sapiens

REFERENCE
1 (bases 1 to 376)

AUTHORS
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.

TITLE
Sequence-tagged connectors: A sequence approach to mapping and

scanning the human genome

JOURNAL
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

MEDLINE
99380589

COMMENT
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3029 row: M column: 8
Class: BAC ends
High quality sequence stop: 376.

FEATURES

Source

BASE COUNT
ORIGIN

131 a 81 c 50 g 114 t

Query Match

Best Local Similarity 63.6%; Score 14; DB 149; Length 376;

Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 tgttgannnnnnnttgata 22

|||||

Db 333 TGTTGGAATTGCGCAATTGATA 312

RESULT 15
AQ844476/c

LOCUS
DEFINITION

an37g11 JM101 filtered library zea mays genomic, DNA sequence.

ACCESSION
AQ844476

VERSION
AQ844476.1 GI:6202964

KEYWORDS
GSS.

SOURCE
Zea mays.

ORGANISM
Zea mays

REFERENCE
1 (bases 1 to 376)

AUTHORS
Rabinowicz,P.D., Schutz,K., Dedhia,N., Yordan,C., Parnell,L.D.,
Stein,L., McCombie,W.R. and Martienssen,R.A.

TITLE
Differential methylation of genes and retrotransposons allows

shotgun sequencing of the maize genome

JOURNAL
Nat. Genet. 23, 305-308 (1999)

COMMENT

Contact: Martienssen RA
Cold Spring Harbor Laboratory
1 Bungtown Rd., Cold Spring Harbor, NY 11724, USA
Tel: 516 367 8322
Fax: 516 367 8369
Email: martienssen@cshl.org
Seq primer: forward
Class: shotgun.

FEATURES

source

Location/Qualifiers

1..376

/organism="Zea mays"

/cultivar="B73"

/db_xref="taxon:4577"

/clone_lib="JM101 filtered library"

/note="Organ: Immature ears; Vector: M13; Site_1: Xba I;

DNA prepared from purified nuclei was digested with the

methylation insensitive enzyme Spe I, size fractionated to

enrich for the 0.5 to 4 kbp fraction, ligated into Xba I

digested M13 vector and electroporated into E.coli JM101.

BASE COUNT 89 a 87 c 100 g 100 t

ORIGIN

Query Match

Best Local Similarity 63.6%; Score 14; DB 163; Length 376;

Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 tgttgannnnnnnttgata 22

|||||

Db 109 TGTTGGATCTGGAAGTTTGATA 88

Search completed: March 27, 2001, 07:38:25

Job time: 4588 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 27, 2001, 08:23:01 ; Search time 207.51 Seconds
(without alignments)
39.827 Million cell updates/sec

Title: US-09-380-826A-7
Perfect score: 22
Sequence: 1 tggtagatcaagatttgata 22

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_36:*

1:	/cgn2_2/gcgdata/geneseq/geneseq/NA1980.DAT:*
2:	/cgn2_2/gcgdata/geneseq/geneseq/NA1981.DAT:*
3:	/cgn2_2/gcgdata/geneseq/geneseq/NA1982.DAT:*
4:	/cgn2_2/gcgdata/geneseq/geneseq/NA1983.DAT:*
5:	/cgn2_2/gcgdata/geneseq/geneseq/NA1984.DAT:*
6:	/cgn2_2/gcgdata/geneseq/geneseq/NA1985.DAT:*
7:	/cgn2_2/gcgdata/geneseq/geneseq/NA1986.DAT:*
8:	/cgn2_2/gcgdata/geneseq/geneseq/NA1987.DAT:*
9:	/cgn2_2/gcgdata/geneseq/geneseq/NA1988.DAT:*
10:	/cgn2_2/gcgdata/geneseq/geneseq/NA1989.DAT:*
11:	/cgn2_2/gcgdata/geneseq/geneseq/NA1990.DAT:*
12:	/cgn2_2/gcgdata/geneseq/geneseq/NA1991.DAT:*
13:	/cgn2_2/gcgdata/geneseq/geneseq/NA1992.DAT:*
14:	/cgn2_2/gcgdata/geneseq/geneseq/NA1993.DAT:*
15:	/cgn2_2/gcgdata/geneseq/geneseq/NA1994.DAT:*
16:	/cgn2_2/gcgdata/geneseq/geneseq/NA1995.DAT:*
17:	/cgn2_2/gcgdata/geneseq/geneseq/NA1996.DAT:*
18:	/cgn2_2/gcgdata/geneseq/geneseq/NA1997.DAT:*
19:	/cgn2_2/gcgdata/geneseq/geneseq/NA1998.DAT:*
20:	/cgn2_2/gcgdata/geneseq/geneseq/NA1999.DAT:*
21:	/cgn2_2/gcgdata/geneseq/geneseq/NA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	22	100.0	22	19 V58901	Leptospi rRNA ge
2	22	100.0	22	19 V58897	L. fairnei nucleoti
3	22	100.0	1477	19 V58896	L. fairnei nucleoti
C 4	16.8	76.4	2582	18 T90543	Maize male fertili
C 5	16.2	73.6	951	20 V84598	Human secreted pro
C 6	16.2	73.6	1291	19 V59636	Human secreted pro
C 7	16.2	73.6	1473	20 X61757	B. burgdorferi ant
8	16.2	73.6	2000	17 T17111	Acetyl-CoA-carboxy
9	16.2	73.6	2001	17 T39904	Maize acetyl CoA c
10	16.2	73.6	2001	21 T49816	ECORI fragment of
11	16.2	73.6	3822	18 T64683	M. leprae gyra pre
12	16.2	73.6	4346	14 QN2933	A3 maize ACCase CD

13	16.2	73.6	7470	17 T39905	Maize acetyl CoA c
14	16.2	73.6	7470	19 V29317	Maize ACCase enzym
15	16.2	73.6	7470	21 Z49820	Maize acetyl CoA c
C 16	16.2	73.6	10811	19 V59091	Potato pollen cell
17	15.8	71.8	843	20 Z15467	Human gene express
18	15.8	71.8	961	19 X14095	H. pylori GHPO 131
19	15.8	71.8	4146	21 Z51556	Human hypoxia resp
C 20	15.6	70.9	297	20 V88753	EST clone HK650.
C 21	15.6	70.9	795	20 V72025	Adenovirus PACTSG2
C 22	15.6	70.9	834	20 V72026	Adenovirus SCAR.RG
C 23	15.6	70.9	1075	18 T72715	C. elegans inhibit
C 24	15.6	70.9	1096	13 Q20498	Encodes fibrinogen
C 25	15.6	70.9	1194	20 V72027	Adenovirus PACSG25
C 26	15.6	70.9	1560	18 T49864	Callosellasma rhod
C 27	15.6	70.9	1584	19 V28845	Human coxsackievir
C 28	15.6	70.9	1761	21 Z45339	DNA encoding a Bac
C 29	15.6	70.9	1989	20 V81481	Black Creek Canal
C 30	15.6	70.9	2354	19 V50429	Human coxsackievir
C 31	15.6	70.9	2851	20 X13369	Enterococcus faeca
C 32	15.6	70.9	7745	18 T79859	Genomic DNA encodi
C 33	15.6	70.9	7745	18 T59308	Sh2-mlRev6 gene (s
C 34	15.6	70.9	7745	20 X04183	Shrunken-2 gene va
C 35	15.6	70.9	9824	14 Q40763	Wild-type shrunken
C 36	15.2	69.1	349	13 Q34809	Human type IV coll
C 37	15.2	69.1	493	20 X13759	Enterococcus faeca
C 38	15.2	69.1	1028	17 T36923	Human OVCA2 tumou
C 39	15.2	69.1	1848	16 Q86996	Aminopeptidase O12
C 40	15.2	69.1	1848	16 Q86998	Aminopeptidase O12
C 41	15.2	69.1	2202	17 T36922	Human OVCA1 tumou
C 42	15.2	69.1	2361	20 X04316	Human secreted pro
C 43	15.2	69.1	2598	18 T72286	Breast cancer mann
C 44	15.2	69.1	2598	21 Z98878	MMTV env protein n
C 45	15.2	69.1	5420	12 Q11643	Partial human comp

ALIGNMENTS

RESULT 1

V58901

ID V58901 standard; DNA; 22 BP.

XX V58901;

XX 20-JAN-1999 (first entry)

XX Leptospi rRNA gene nucleotide sequence.

XX Infection; pathogenic Leptospi; protective immunity; therapy; diagnosis; ss.

XX Leptospi sp.

XX WO9840099-A1.

XX 17-SEP-1998.

XX 06-MAR-1998; 98WO-AU00145.

XX 07-MAR-1997; 97AU-0005494.

XX (AGRI-) AGRIC VICTORIA SERVICES PTY LTD.

XX (PIGR-) PIG RES & DEV CORP.

XX Chappel RJ;

XX WPI; 1998-520791/44.

XX New isolated pathogenic Leptospi bacterium - useful for, e.g

XX developing products for conferring protective immunity, and for

XX prophylactic or therapeutic treatment

XX Claim 15; Page 72; 94pp; English.

XX This sequence represents a Leptospira DNA sequence isolated from the
 CC pathogenic Leptospira (LS) bacterium of the invention. The bacterium
 CC belongs to serogroup Hurstbridge or serovar hurstbridge or the species
 CC L. fainei. The LS bacteria can be used for conferring protective
 CC immunity against pathogenic LS bacteria in humans or animals. The
 CC bacteria can also be used for prophylactic or therapeutic treatment of LS
 CC infections. The DNAs and antibodies may also be used for detection and
 CC diagnosis of past or present LS infection.
 XX
 SQ Sequence 22 BP; 7 A; 2 C; 5 G; 8 T; 0 other;

Query Match 100.0%; Score 22; DB 19; Length 22;
 Best Local Similarity 100.0%; Pred. No. 0.047;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgttgatcacagaattgata 22
 |||||
 Db 1 tgttgatcacagaattgata 22

RESULT 2

V58897
 ID V58897 standard; DNA; 22 BP.

AC V58897;

XX 20-JAN-1999 (first entry)

DE L. fainei nucleotide sequence.

XX Infection; pathogenic Leptospira; protective immunity; therapy;
 KW diagnosis; ss.

XX Leptospira fainei.

XX WO9840099-A1.

XX 17-SEP-1998.

XX 06-MAR-1998; 98WO-AU00145.

XX 07-MAR-1997; 97AU-0005494.

XX (AGRI-) AGRIC VICTORIA SERVICES PTY LTD.

XX (PIGR-) PIG RES & DEV CORP.

XX Chappel RJ;

XX WPI; 1998-520791/44.

XX New isolated pathogenic Leptospira bacterium - useful for, e.g
 PT developing products for conferring protective immunity, and for
 PT prophylactic or therapeutic treatment

XX Claim 15; Page 70; 94pp; English.

XX This sequence represents a Leptospira DNA sequence isolated from the
 CC pathogenic Leptospira (LS) bacterium of the invention. The bacterium
 CC belongs to serogroup Hurstbridge or serovar hurstbridge or the species
 CC L. fainei. The LS bacteria can be used for conferring protective
 CC immunity against pathogenic LS bacteria in humans or animals. The
 CC bacteria can also be used for prophylactic or therapeutic treatment of LS
 CC infections. The DNAs and antibodies may also be used for detection and
 CC diagnosis of past or present LS infection.
 XX

SQ Sequence 22 BP; 7 A; 2 C; 5 G; 8 T; 0 other;

Query Match 100.0%; Score 22; DB 19; Length 22;
 Best Local Similarity 100.0%; Pred. No. 0.047;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgttgatcacagaattgata 22
 |||||
 Db 1 tgttgatcacagaattgata 22

RESULT 3

V58896

ID V58896 standard; DNA; 1477 BP.

XX V58896;

XX 20-JAN-1999 (first entry)

DE L. fainei nucleotide sequence.

XX Infection; pathogenic Leptospira; protective immunity; therapy;
 KW diagnosis; ss.

XX Leptospira fainei.

XX WO9840099-A1.

XX 17-SEP-1998.

XX 06-MAR-1998; 98WO-AU00145.

XX 07-MAR-1997; 97AU-0005494.

XX (AGRI-) AGRIC VICTORIA SERVICES PTY LTD.

XX (PIGR-) PIG RES & DEV CORP.

XX Chappel RJ;

XX WPI; 1998-520791/44.

XX New isolated pathogenic Leptospira bacterium - useful for, e.g
 PT developing products for conferring protective immunity, and for
 PT prophylactic or therapeutic treatment

XX Claim 15; Page 69-70; 94pp; English.

XX This sequence represents a Leptospira DNA sequence isolated from the
 CC pathogenic Leptospira (LS) bacterium of the invention. The bacterium
 CC belongs to serogroup Hurstbridge or serovar hurstbridge or the species
 CC L. fainei. The LS bacteria can be used for conferring protective
 CC immunity against pathogenic LS bacteria in humans or animals. The
 CC bacteria can also be used for prophylactic or therapeutic treatment of LS
 CC infections. The DNAs and antibodies may also be used for detection and
 CC diagnosis of past or present LS infection.
 XX

SQ Sequence 1477 BP; 390 A; 334 C; 439 G; 314 T; 0 other;

Query Match 100.0%; Score 22; DB 19; Length 1477;
 Best Local Similarity 100.0%; Pred. No. 0.072;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgttgatcacagaattgata 22
 |||||
 Db 154 tgttgatcacagaattgata 175

RESULT 4

T90543/C

ID T90543 standard; DNA; 2582 BP.

XX T90543;

XX 13-FEB-1998 (first entry)

XX Maize male fertility gene 233 Zm41-A.

XX

KW Zm41-A; male fertility; hybrid seed; transgenic plant; antisense;
 KW ribozyme; male sterile; maize; Ms41-A; ds.
 XX
 OS
 XX
 FH Location/Qualifiers
 FT exon 823..915
 FT /tag= a
 FT intron 916..935
 FT /tag= b
 FT exon 936..1090
 FT /tag= c
 FT intron 1091..1167
 FT /tag= d
 FT exon 1168..1332
 FT /tag= e
 XX
 XX WO9723618-A1.
 PN
 XX
 PD 03-JUL-1997.
 XX
 XX 20-DEC-1996; 96WO-GB03191.
 PF
 XX 21-DEC-1995; 95GB-0026218.
 PR
 XX (GENE-) GENE SHEARS PTY LTD.
 PA
 XX
 XX Baudot G, Garcia D, Hodge R, Perez P;
 PI
 XX WPI; 1997-351055/32.
 DR
 XX
 XX Nucleic acid encoding proteins involved in male fertility in plants
 PT - used to control fertility and for production of hybrid seed
 PT
 XX
 PS Claim 3; Fig 15; 85pp; English.
 XX
 CC This DNA sequence comprises the maize Z33 Zm41-A gene. This is an
 CC orthologue of Arabidopsis Ms41-A (see T90522), a gene that confers
 CC male fertility. It was obtained following a database search for
 CC sequences that showed homology to Ms41-A DNA, isolate of a
 CC partial cDNA clone, and use of this clone to isolate Zm41-A genes
 CC Z31 (T90542), Z33 (T90543) and Z35 (T90544) from genomic lambda
 CC libraries. The genes show a high level of conservation. Z35 may
 CC be derived from Z31 via genetic rearrangements, deletions and/or
 CC insertions. Z33 has subsequent deletions from Z35 and is
 CC truncated, having only exons 3, 5 and 6. The Ms41-A and Zm41-A
 CC genes, antisense or ribozyme sequences can be used to produce
 CC transgenic plants with controlled male fertility. Male sterile
 CC plants are useful for hybrid seed production, particularly in
 CC plants where restoration of fertility is not needed, e.g.
 CC Brassicaceae, lettuce, spinach and onions.
 CC
 XX Sequence 2582 BP: 670 A; 512 C; 537 G; 863 T; 0 other;
 SQ
 Query Match 76.4%; Score 16.8; DB 18; Length 2582;
 Best Local Similarity 90.0%; Pred. No. 26;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 OY 1 tgttgatcacaaagttaga 20
 ||||| ||||| ||||| |||||
 Db 268 TGTTGAATCACAGACTGA 249
 RESULT 5
 V84598/c
 ID V84598 standard; DNA; 951 BP.
 XX
 AC V84598;
 XX
 XX 01-MAR-1999 (first entry)
 DT
 XX Human secreted protein gene 188 clone HHP8F70.
 DE

XX Human; secreted protein; fusion protein; gene therapy; protein therapy;
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
 KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
 KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
 KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
 XX
 OS Homo sapiens.
 XX
 XX WO9854963-A2.
 PN
 XX
 PD 10-DEC-1998.
 XX
 XX 04-JUN-1998; 98WO-US11422.
 PF
 XX
 XX 18-DEC-1997; 97US-0070923.
 PR 06-JUN-1997; 97US-0048877.
 PR 06-JUN-1997; 97US-0048881.
 PR 06-JUN-1997; 97US-0048884.
 PR 06-JUN-1997; 97US-0048893.
 PR 06-JUN-1997; 97US-0048896.
 PR 06-JUN-1997; 97US-0048899.
 PR 06-JUN-1997; 97US-0048915.
 PR 06-JUN-1997; 97US-0048949.
 PR 06-JUN-1997; 97US-0048964.
 PR 06-JUN-1997; 97US-0048972.
 PR 06-JUN-1997; 97US-0049020.
 PR 06-JUN-1997; 97US-0049375.
 PR 05-SEP-1997; 97US-0057628.
 PR 05-SEP-1997; 97US-0057635.
 PR 05-SEP-1997; 97US-0057644.
 PR 05-SEP-1997; 97US-0057647.
 PR 05-SEP-1997; 97US-0057650.
 PR 05-SEP-1997; 97US-0057661.
 PR 05-SEP-1997; 97US-0057667.
 PR 05-SEP-1997; 97US-0057761.
 PR 05-SEP-1997; 97US-0057764.
 PR 05-SEP-1997; 97US-0057770.
 PR 05-SEP-1997; 97US-0057775.
 PR 05-SEP-1997; 97US-0057778.
 PR 06-JUN-1997; 97US-0048875.
 PR 06-JUN-1997; 97US-0048878.
 PR 06-JUN-1997; 97US-0048882.
 PR 06-JUN-1997; 97US-0048885.
 PR 06-JUN-1997; 97US-0048894.
 PR 06-JUN-1997; 97US-0048897.
 PR 06-JUN-1997; 97US-0048900.
 PR 06-JUN-1997; 97US-0048916.
 PR 06-JUN-1997; 97US-0048962.
 PR 06-JUN-1997; 97US-0048970.
 PR 06-JUN-1997; 97US-0048974.
 PR 06-JUN-1997; 97US-0049373.
 PR 05-SEP-1997; 97US-0057584.
 PR 05-SEP-1997; 97US-0057629.
 PR 05-SEP-1997; 97US-0057642.
 PR 05-SEP-1997; 97US-0057645.
 PR 05-SEP-1997; 97US-0057648.
 PR 05-SEP-1997; 97US-0057651.
 PR 05-SEP-1997; 97US-0057662.
 PR 05-SEP-1997; 97US-0057668.
 PR 05-SEP-1997; 97US-0057762.
 PR 05-SEP-1997; 97US-0057765.
 PR 05-SEP-1997; 97US-0057771.
 PR 06-JUN-1997; 97US-0057776.
 PR 06-JUN-1997; 97US-0048876.
 PR 06-JUN-1997; 97US-0048880.
 PR 06-JUN-1997; 97US-0048883.
 PR 06-JUN-1997; 97US-0048892.
 PR 06-JUN-1997; 97US-0048895.
 PR 06-JUN-1997; 97US-0048898.
 PR

PR 06-JUN-1997; 97US-0048901.
 PR 06-JUN-1997; 97US-0048917.
 PR 06-JUN-1997; 97US-0048963.
 PR 06-JUN-1997; 97US-0048971.
 PR 06-JUN-1997; 97US-0049019.
 PR 06-JUN-1997; 97US-0049374.
 PR 06-JUN-1997; 97US-0057627.
 PR 05-SEP-1997; 97US-0057634.
 PR 05-SEP-1997; 97US-0057634.
 PR 05-SEP-1997; 97US-0057643.
 PR 05-SEP-1997; 97US-0057646.
 PR 05-SEP-1997; 97US-0057649.
 PR 05-SEP-1997; 97US-0057654.
 PR 05-SEP-1997; 97US-0057666.
 PR 05-SEP-1997; 97US-0057760.
 PR 05-SEP-1997; 97US-0057763.
 PR 05-SEP-1997; 97US-0057769.
 PR 05-SEP-1997; 97US-0057774.
 PR 05-SEP-1997; 97US-0057777.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Brewer LA, Carter KC, Dillon PJ, Ebner R, Endress GA;
 PI Fan P, Feng P, Ferrie AM, Fischer CL, Florence C;
 PI Florence K, Greene JM, Hu J, Kyaw H, Lafleur DW;
 PI Li Y, Moore PA, Ni J, Olsen HS, Rosen CA, Ruben SM;
 PI Shi Y, Soppet DR, Wei Y, Young P, Yu G, Zeng Z;
 XX
 WI 1999-059865/05.
 DR P-PSDB; W88721.
 DR
 XX
 PT New isolated human genes and the secreted polypeptides they encode -
 PT useful for diagnosis and treatment of e.g. cancers, neurological
 PT disorders, immune diseases, inflammation or blood disorders
 XX
 PS Claim 4; Page 450-451; 772pp; English.
 XX
 CC The invention relates to nucleic acid sequences (V84411 to V84633)
 CC encoding human secreted proteins (W88534 to W88756). The secreted protein
 CC gene sequences are deposited with the ATCC under deposit numbers ATCC
 CC 97979, 97974, 97975, 97976, 97977, 209007, 209008, 209009, 209010,
 CC 209011, 209080, 209081, 209082, 209083, 209084, 209085, 209511. Host
 CC cells comprising recombinant vectors containing the nucleic acid
 CC sequences are used for the recombinant production of the secreted
 CC proteins. The polynucleotide and amino acid sequences are useful for are
 CC useful for preventing, treating or ameliorating medical conditions e.g.
 CC by protein or gene therapy. Pathological conditions can be also diagnosed
 CC by determining the amount of the new polypeptides in a sample or by
 CC determining the presence of mutations in the new polynucleotides.
 CC Specific uses are described for each of the polynucleotides, based on
 CC which tissues they are most highly expressed in, and include developing
 CC products for the diagnosis or treatment of cancer, neurodegenerative
 CC disorders, developmental abnormalities and foetal deficiencies, blood
 CC disorders, tumours, leukemias, diseases of the immune system, autoimmune
 CC diseases, hepatic and renal disease, lymphomas, inflammation, allergies,
 CC ischemic shock, Alzheimer's and cognitive disorders, schizophrenia,
 CC restenosis, prostate diseases, obesity, disorders involving osteoclasts
 CC such as osteoporosis, arthritis or malignancies, diseases of testes, lung
 CC or thymus, digestive/endocrine disorders, infections and AIDS. The
 CC polypeptides are also useful for identifying their binding partners.
 CC The present sequence represents a gene encoding a human secreted protein
 CC (see descriptor line for gene number and clone identification).
 XX
 SQ Sequence 951 BP; 296 A; 134 C; 192 G; 329 T; 0 other;

RESULT 6
 V59636/C
 ID V59636 standard; DNA; 1291 BP.
 XX
 AC V59636;
 XX
 DT 19-JAN-1999 (first entry)
 XX
 DE Human secreted protein gene 126 clone HELGH31.
 XX
 KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
 KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
 KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
 KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
 XX
 OS Homo sapiens.
 PN W09839448-A2.
 XX
 PD 11-SEP-1998.
 XX
 PF 06-MAR-1998; 98WO-US04493.
 XX
 PR 02-OCT-1997; 97US-0061060.
 PR 07-MAR-1997; 97US-0038621.
 PR 07-MAR-1997; 97US-0040161.
 PR 07-MAR-1997; 97US-0040162.
 PR 07-MAR-1997; 97US-0040163.
 PR 07-MAR-1997; 97US-0040333.
 PR 07-MAR-1997; 97US-0040334.
 PR 07-MAR-1997; 97US-0040336.
 PR 07-MAR-1997; 97US-0040626.
 PR 11-APR-1997; 97US-0043311.
 PR 11-APR-1997; 97US-0043312.
 PR 11-APR-1997; 97US-0043313.
 PR 11-APR-1997; 97US-0043314.
 PR 11-APR-1997; 97US-0043568.
 PR 11-APR-1997; 97US-0043569.
 PR 11-APR-1997; 97US-0043576.
 PR 11-APR-1997; 97US-0043578.
 PR 11-APR-1997; 97US-0043580.
 PR 11-APR-1997; 97US-0043669.
 PR 11-APR-1997; 97US-0043670.
 PR 11-APR-1997; 97US-0043671.
 PR 11-APR-1997; 97US-0043672.
 PR 11-APR-1997; 97US-0043674.
 PR 23-MAY-1997; 97US-0047492.
 PR 23-MAY-1997; 97US-0047500.
 PR 23-MAY-1997; 97US-0047501.
 PR 23-MAY-1997; 97US-0047502.
 PR 23-MAY-1997; 97US-0047503.
 PR 23-MAY-1997; 97US-0047581.
 PR 23-MAY-1997; 97US-0047582.
 PR 23-MAY-1997; 97US-0047583.
 PR 23-MAY-1997; 97US-0047584.
 PR 23-MAY-1997; 97US-0047585.
 PR 23-MAY-1997; 97US-0047586.
 PR 23-MAY-1997; 97US-0047587.
 PR 23-MAY-1997; 97US-0047588.
 PR 23-MAY-1997; 97US-0047589.
 PR 23-MAY-1997; 97US-0047590.
 PR 23-MAY-1997; 97US-0047592.
 PR 23-MAY-1997; 97US-0047593.
 PR 23-MAY-1997; 97US-0047594.
 PR 23-MAY-1997; 97US-0047595.
 PR 23-MAY-1997; 97US-0047596.
 PR 23-MAY-1997; 97US-0047597.
 PR 23-MAY-1997; 97US-0047598.
 PR 23-MAY-1997; 97US-0047599.

Query Match 73.6%; Score 16.2; DB 20; Length 951;
 Best Local Similarity 85.7%; Pred. No. 47;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 tggtagatcacaaatttgat 21
 Db 524 tggtagatcacaaatttgat 504

23-MAY-1997; 97US-0047599.
 23-MAY-1997; 97US-0047600.
 23-MAY-1997; 97US-0047601.
 23-MAY-1997; 97US-0047612.
 23-MAY-1997; 97US-0047613.
 23-MAY-1997; 97US-0047614.
 23-MAY-1997; 97US-0047615.
 23-MAY-1997; 97US-0047617.
 23-MAY-1997; 97US-0047618.
 23-MAY-1997; 97US-0047632.
 23-MAY-1997; 97US-0047633.
 06-JUN-1997; 97US-0048964.
 06-JUN-1997; 97US-0048974.
 13-JUN-1997; 97US-0049610.
 08-JUL-1997; 97US-0051926.
 16-JUL-1997; 97US-0052874.
 18-AUG-1997; 97US-0055724.
 22-AUG-1997; 97US-0056630.
 22-AUG-1997; 97US-0056631.
 22-AUG-1997; 97US-0056632.
 22-AUG-1997; 97US-0056636.
 22-AUG-1997; 97US-0056637.
 22-AUG-1997; 97US-0056662.
 22-AUG-1997; 97US-0056664.
 22-AUG-1997; 97US-0056845.
 22-AUG-1997; 97US-0056862.
 22-AUG-1997; 97US-0056864.
 22-AUG-1997; 97US-0056872.
 22-AUG-1997; 97US-0056874.
 22-AUG-1997; 97US-0056875.
 22-AUG-1997; 97US-0056876.
 22-AUG-1997; 97US-0056877.
 22-AUG-1997; 97US-0056878.
 22-AUG-1997; 97US-0056879.
 22-AUG-1997; 97US-0056880.
 22-AUG-1997; 97US-0056881.
 22-AUG-1997; 97US-0056882.
 22-AUG-1997; 97US-0056884.
 22-AUG-1997; 97US-0056886.
 22-AUG-1997; 97US-0056887.
 22-AUG-1997; 97US-0056888.
 22-AUG-1997; 97US-0056889.
 22-AUG-1997; 97US-0056892.
 22-AUG-1997; 97US-0056893.
 22-AUG-1997; 97US-0056894.
 22-AUG-1997; 97US-0056903.
 22-AUG-1997; 97US-0056908.
 22-AUG-1997; 97US-0056909.
 22-AUG-1997; 97US-0056910.
 22-AUG-1997; 97US-0056911.
 05-SEP-1997; 97US-0057650.
 05-SEP-1997; 97US-0057669.
 05-SEP-1997; 97US-0057761.
 12-SEP-1997; 97US-0058785.
 (HUMA-) HUMAN GENOME SCI INC.
 Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA,
 Feng P, Ferrle AM, Fischer CL, Florence KA, Greene JM, Hu JS,
 Kyaw H, Lafleur DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA,
 Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;
 WPI; 1998-506364/43.
 P-PSDB; W74854.
 New isolated human genes and the secreted polypeptide(s) they encode
 - useful for diagnosis and treatment of e.g. cancers, neurological
 disorders, immune diseases, inflammation or blood disorders
 Claim 1; Page 360-361; 721pp; English.
 This sequence represents a nucleic acid molecule designated Gene 126 from
 the human cDNA clone HELGH31 (deposited as clone ATCC 97902 and ATCC

CC 209048) which encodes a secreted human protein. The gene can be used to
 CC generate fusion proteins by linking to the gene to a human immunoglobulin
 CC Fc portion (e.g. V59502) for increasing the stability of the fused
 CC protein as compared to the human protein only.
 CC The invention relates to 186 novel genes and their fragments (nucleic
 CC acid sequences: V59511-V59812; amino acid sequences W74731-W75026) which
 CC are useful for preventing, treating or ameliorating medical conditions
 CC e.g. by protein or gene therapy. Also, pathological conditions can be
 CC diagnosed by determining the amount of the new polypeptides in a sample
 CC or by determining the presence of mutations in the new polynucleotides.
 CC Specific uses are described for each of the 186 polynucleotides, based on
 CC which tissues they are most highly expressed in (see V59511 for described
 CC uses).
 XX Sequence 1291 BP; 391 A; 228 C; 203 G; 469 T; 0 other;
 SQ
 Query Match 73.6%; Score 16.2; DB 19; Length 1291;
 Best Local Similarity 85.7%; Pred. No. 48;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 tgttgatcacagatttgat 21
 Db 367 TGTATGATCACATGATTGAT 347
 RESULT 7
 X61757/c
 ID X61757 standard; DNA; 1473 BP.
 XX
 AC X61757;
 XX
 DT 19-JUL-1999 (first entry)
 XX
 DE B. burgdorferi antigenic protein coding sequence, f893.nt.
 XX
 KW Antigenic protein; vaccine; Lyme disease; infection; detection; ss.
 XX
 OS Borrelia burgdorferi.
 XX
 PN WO9859071-A1.
 XX
 PD 30-DEC-1998.
 XX
 PF 18-JUN-1998; 98WO-0512718.
 XX
 PR 03-SEP-1997; 97US-0057483.
 PR 20-JUN-1997; 97US-0050359.
 PR 22-JUL-1997; 97US-0053344.
 PR 22-JUL-1997; 97US-0053377.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (MEDI-) MEDIMMUNE INC.
 XX
 PI Choi GH, Erwin AL, Hanson MS, Lathigra R;
 XX
 DR WPI; 1999-189980/16.
 DR P-PSDB; Y20060.
 XX
 PT New isolated Borrelia burgdorferi nucleic acids - used to develop
 PT products for the diagnosis, prevention and treatment of diseases
 PT caused by Borrelia, particularly Lyme disease
 XX
 PS Claim 1; Page 180; 275pp; English.
 XX
 CC This sequence encodes a Borrelia burgdorferi (Bb) protein of the
 CC invention, which is suitable for use in a vaccine. The Bb polypeptides
 CC can be used in vaccines for eliciting protective antibodies to members of
 CC the Borrelia genus, particularly for the use against Lyme disease in
 CC humans and animals. They can be used for preventing or attenuating an
 CC infection caused by a member of the Borrelia genus. The products can also
 CC be used for detection of members of the Borrelia genus.

0:

QY 2 gttggatcacagaatttgata 22
 ID T64683 standard; DNA; 2001 BP.
 DB 42 gttggatcacagaatttgata 62

RESULT 10
 249816
 ID 249816 standard; DNA; 2001 BP.
 XX AC 249816;
 XX DE
 DT 18-APR-2000 (first entry)
 DE EcoRI fragment of lambda clone #15-14 with maize ACCase gene portion.
 KW Herbicide resistance; gene modification; lambda clone #15-14;
 KW maize acetyl CoA carboxylase; ACCase; herbicide; plant oil content;
 KW marker-assisted plant selection; groat oil trait;
 KW restriction fragment length polymorphism mapping;
 KW high-energy animal feed; low-fat human food; ds.
 XX Zea mays.
 OS
 XX
 XX WO9967367-Al.
 PN
 XX
 XX 29-DEC-1999.
 XX
 XX 22-JUN-1999; 99WO-US14022.
 XX
 XX 22-JUN-1998; 98US-0090240.
 PR
 XX 02-JUL-1998; 98US-0091640.
 XX
 XX (MINU) UNIV MINNESOTA.
 PA (USDA) US DEPT OF AGRICULTURE.
 PA (EGLI/) EGLI M A.
 PA (GROH/) GROH S.
 PA (KIAN/) KIANIAN S F.
 PA (PHIL/) PHILLIPS R L.
 PA (RINE/) RINES H W.
 PA (SOME/) SOMERS D A.
 XX
 XX Egli MA, Groh S, Kianian SF, Phillips RL, Rines HW, Somers DA;
 PI WPI: 2000-147205/13.
 XX
 XX New DNA encoding acetyl-CoA carboxylase from oats, used to produce
 PT transformed plants with herbicide resistance and altered oil content
 PT
 XX Disclosure; Fig 10; 197pp; English.
 XX
 XX The present DNA sequence is a 2kb EcoRI fragment of lambda clone #15-14
 CC including a portion of a maize acetyl CoA carboxylase (ACCase) gene
 CC located at bases 2883 to 83 from the 3' stop codon. ACCase is involved
 CC in fatty acid synthesis and is the target of the specified herbicides.
 CC Transformation of plants with ACCase imparts resistance to
 CC cyclohexanedione and aryloxyphenoxypropionic acid herbicides and alter
 CC the oil content. The ACCase DNA is also used as source of probes and
 CC primers for the identification of transgenic plants; in marker-assisted
 CC plant selection and for restriction fragment length polymorphism
 CC mapping, used for high-energy animal feed and high-fiber, low-fat human
 CC food and in genetic dissection of the groat oil trait.
 XX
 XX Sequence 2001 BP; 546 A; 406 C; 494 G; 554 T; 1 other;

Query Match 73.6%; Score 16.2; DB 21; Length 2001;
 Best Local Similarity 85.7%; Pred. No. 51;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 gttggatcacagaatttgata 22
 ID T64683 standard; DNA; 2001 BP.
 DB 42 gttggatcacagaatttgata 62

RESULT 11
 T64683
 ID T64683 standard; DNA; 3822 BP.
 XX AC T64683;
 XX DT 17-MAR-1998 (first entry)
 XX DE
 XX M. leprae gyrA precursor coding sequence.
 XX Mycobacterium sp.; internal sequence; intein; immature; gyrase;
 KW protein splicing; precursor; gyrA; ss.
 KW Mycobacterium leprae.
 OS
 XX FR2739859-Al.
 PN
 XX 18-APR-1997.
 PD
 XX 17-OCT-1995; 95PR-0012162.
 PF
 XX 17-OCT-1995; 95PR-0012162.
 PR
 XX (INSP) INST PASTEUR.
 PA
 XX Cole S, Fsihi H;
 PI WPI: 1997-247976/23.
 DR P-PSDB; W15078.
 XX
 XX Mycobacterial DNA gyrase precursor protein - and polypeptide(s)
 PT corresponding to mycobacterial DNA gyrase intein sequences
 PT
 XX Claim 5; Pages 33-34; 55pp; French.
 XX
 XX This DNA, isolated from Mycobacterium leprae, encodes a precursor
 CC (immature) gyrase protein from which an "intein" (see W15074) is excised
 CC during maturation. The gyrase and its coding sequence are useful to
 CC fabricate products that alter the maturation of proteins essential for
 CC the development of infectious agents by altering the protein splicing of
 CC precursor polypeptides of the proteins. The inteins encoded by the gyrA
 CC genes of Mycobacterium leprae, M. flavesens, M. goodii and M. kansasii
 CC are used: (a) to modify the genome of a eukaryotic cell that lacks
 CC endogenous biological activity identical to that of the polypeptide;
 CC (b) to replace a copy of a gene present in a recipient genome by
 CC integration of a gene different from that where the integration takes
 CC place; and (c) for targeted insertion of a foreign DNA sequence into a
 CC selected site in the genome of a eukaryotic cell not containing the
 CC specific cleavage sites of the polypeptide.
 XX
 XX Sequence 3822 BP; 843 A; 899 C; 1092 G; 988 T; 0 other;

Query Match 73.6%; Score 16.2; DB 18; Length 3822;
 Best Local Similarity 85.7%; Pred. No. 54;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 gttggatcacagaatttgata 22
 ID T64683 standard; DNA; 3822 BP.
 DB 2334 gttggatcacagaatttgata 2354

RESULT 12
 Q42933
 ID Q42933 standard; DNA; 4346 BP.
 XX AC Q42933;
 XX DT 15-OCT-1993 (first entry)
 XX A3 maize ACCase cDNA clone.

XX Acetyl CoA carboxylase; ACCase; probe; expression; oil content;
 KW lower; modified; increased; oilseed rape; sunflower; resistant;
 KW resistance; grass-weed herbicides; fluzafop; alkylketones; maize;
 KW wheat; barley; ss.
 XX
 OS Zea mays.
 XX
 FH Key Location/Qualifiers
 FT CDS 3..3947
 FT /*tag= a
 FT /note= "ACCase"
 FT misc_feature 1875..1878
 FT /*tag= b
 FT /note= "GAN encodes Glu or Asp"
 XX
 PN W09311243-A.
 XX
 XX
 PD 10-JUN-1993.
 XX
 XX 27-NOV-1992; 92WO-GB02205.
 XX
 XX 28-NOV-1991; 91GB-0025330.
 XX
 PA (ICIL) ICI AUSTRALIA OPERATIONS PTY LTD.
 PA (ICIL) IMPERIAL CHEM IND PLC.
 XX
 XX Ashton AR, Jenkins CL, Whitfield PR;
 XX
 DR WPI: 1993-197061/24.
 DR P-PSDB; R36781.
 XX
 XX DNA clones for use in probing plant DNA - comprise maize acetyl
 PT coenzyme-A carboxylase gene flanked by heterologous DNA
 PT
 XX
 PS Claim 1; Fig 3; 69pp; English.
 XX
 CC The sequence is that of the A3 acetyl CoA carboxylase (ACCase) cDNA
 CC maize clone which can be used to probe plant DNA to isolate other
 CC such clones. It may be used with plant regulatory sequences to
 CC produce expression cassettes for ACCase. These cassettes are used to
 CC transform plants to downregulate prodn. of ACCase. This can alter
 CC the composition of seeds or other plant parts, e.g. enabling oil-
 CC bearing plants such as oilseed rape, sunflower or oilpalm, having a
 CC lower or modified oil content to be produced. It can be used to form
 CC expression cassettes for overexpression of ACCase., leading to prodn.
 CC of plants with an increased oil content. It may also be used to recover
 CC the ACCase gene promoter. This can be used to generate RNA in a
 CC tissue-specific or developmentally regulated manner, and this RNA
 CC used to inhibit ACCase expression. Monocotyledonous plants can be made
 CC resistant to grass-weed herbicides such as fluzafop and alkylketones
 CC by transforming with cassettes adapted to express ACCase. This is
 CC achieved by overexpression of monocot ACCase, expression of dicot
 CC ACCase (which is relatively tolerant to these herbicides), or
 CC possibly by expression of a resistant form of maize ACCase. This
 CC provides crop plants which are resistant to herbicides used to
 CC destroy weeds growing between the plants. This allows the overall
 CC applicn. of herbicide without effecting the crop, and is also useful
 CC where there has been short term carryover of herbicide from the
 CC previous crop. Using the method, weeds such as wild oats may be
 CC controlled, and the farmer is given extra options to obtain an improved
 CC harvest using means which are safer, cheaper or more effective.
 CC Suitable crops are maize, wheat and barley.
 XX
 SQ Sequence 4346 BP; 1239 A; 829 C; 1064 G; 1213 T; 1 other;

Query Match 73.6%; Score 16.2; DB 14; Length 4346;
 Best Local Similarity 85.7%; Pred. No. 55;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 2 gttggatcacaaagtattgata 22
 ||||| ||||| ||| ||

Db 902 gttggatgacaagagattgtta 922
 RESULT 13
 T39905
 ID T39905 standard; cDNA; 7470 BP.
 XX
 AC T39905;
 XX
 DT 23-JAN-1997 (first entry)
 XX
 DE Maize acetyl CoA carboxylase cDNA.
 XX
 KW Acetyl CoA carboxylase; ACCase; herbicide tolerance;
 KW cyclohexanedione; aryloxyphenoxypionic acid; vegetable oil;
 KW oilseed; maize; corn; ss.
 XX
 OS Zea mays line A188.
 XX
 FH Key Location/Qualifiers
 FT CDS 37..7014
 FT /*tag= a
 FT /EC_number= 6.4.1.2
 XX
 PN W09631609-A2.
 XX
 XX 10-OCT-1996.
 PD
 XX
 XX 04-APR-1996; 96WO-US04625.
 PF
 XX
 XX 05-APR-1995; 95US-0417089.
 PR
 XX
 XX (MINU) UNIV MINNESOTA.
 PA
 XX Egli MA, Gengenbach BG, Gronwald JW, Lutz SM, Somers DA;
 PI Wyse DL;
 XX
 DR WPI; 1996-465030/46.
 DR P-PSDB; W05590.
 XX
 XX DNA encoding maize acetyl coenzyme A carboxylase gene - used for
 PT prodn. of plants with herbicide tolerance or altered oil content
 PT
 XX Claim 2; Page 78-80; 131pp; English.
 PS
 XX A cDNA sequence (T39905) codes for maize acetyl CoA carboxylase
 CC (ACCase) (W05590), an enzyme that plays a central role in fatty
 CC acid biosynthesis and accumulation in plants and seeds. The
 CC complete sequence was deduced from cDNA clone #18-5, obt'd. from
 CC a lambda gt10 cDNA library of maize inbred A188 seedlings, and from
 CC PCR amplifications (see also T39918-23). Genomic ACCase DNA
 CC fragments have also been isolated (see also T39906-16). ACCase DNA
 CC can be incorporated into a vector and used to increase the
 CC herbicide tolerance or oil content of a transgenic plant, or used
 CC to produce ACCase in heterologous systems.
 XX
 SQ Sequence 7470 BP; 2119 A; 1429 C; 1844 G; 2071 T; 7 other;
 Query Match 73.6%; Score 16.2; DB 17; Length 7470;
 Best Local Similarity 85.7%; Pred. No. 58;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 2 gttggatcacaaagtattgata 22
 ||||| ||||| ||| ||
 Db 3972 gttggatgacaagagattgtta 3992
 RESULT 14
 V29317
 ID V29317 standard; cDNA; 7470 BP.
 XX
 AC V29317;

```
XX 30-JUL-1998 (first entry)
XX Maize ACCase enzyme encoding cDNA.
XX ACCase; maize; herbicide resistant; corn plant; tolerance: Acc1; Acc2;
XX cyclohexanedione; CHD; aryloxyphenoxypyranoic acid; APA; enzyme; ss.
XX Zea mays.
XX Key Location/Qualifiers
XX CDS 37..7014
XX /*tag= a
XX /transl_except= (pos:4339..4341, aa: His)
XX /transl_except= (pos:4456..4458, aa: Ser)
XX /transl_except= (pos:4359..4461, aa: Lys)
XX /product= "ACCase"
XX W09808963-A1.
XX 05-MAR-1998.
XX 29-AUG-1997; 97WO-US15344.
XX 30-AUG-1996; 96US-0697826.
XX (MINU ) UNIV MINNESOTA.
XX Egli MA, Gengenbach BG, Lutz SM, Marshall LC, Parker WB;
XX Somers DA, Vandee KL, Wyse DL;
XX WPI; 1998-207043/18.
XX P-PSDB; W56736.
XX Herbicide resistant corn plants - prepared using Acc1 and Acc2 gene
XX combinations
XX Example 7; Fig 3A-E; 112pp; English.
XX This cDNA encodes a maize ACCase enzyme. This can be used in a method
XX of preparing an herbicide resistant corn plant which comprises crossing a
XX first corn plant to a second corn plant so as to yield a progeny plant,
XX where the first corn plant comprises at least 1 herbicide resistant
XX allele and the second plant comprises at least 1 herbicide resistant
XX allele which is not allelic to the herbicide resistant allele present in
XX the first plant. The herbicide resistant corn plants are prepared using
XX Acc1 and Acc2 gene combinations which impart cyclohexanedione (CHD) or
XX aryloxyphenoxypyranoic acid (APA) herbicide tolerance to the corn plant.
XX The methods are used to impart CHD and APA herbicide tolerance to corn
XX plants and to produce CHD or APA herbicide tolerant Zea mays (corn)
XX homozygous or heterozygous for Acc1 and homozygous or heterozygous for
XX Acc2. The methods can also be used to impart tolerance to a corn plant
XX to an agent which inhibits acetyl CoA carboxylase, selected from
XX 3-(2,4-dichlorophenyl)-perhydroindolizine-2,4-dione, 3-isopropyl-6-
XX (N-[2,2-dimethylpropyl]-acetamido-1,3,5-triazine-2,4 (1H,3H)dione,
XX soraphen A and their structural analogues.
XX Sequence 7470 BP; 2119 A; 1429 C; 1844 G; 2071 T; 7 other;

Query Match 73.6%; Score 16.2; DB 19; Length 7470;
Best Local Similarity 85.7%; Pred. No. 58;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 gttggatcacaaagatttgata 22
Db 3972 gttggatgacaaagatttgata 3992
||||| ||||| |||||

RESULT 15
249820
ID 249820 standard; cDNA; 7470 BP.
XX
```

```
AC 249820;
XX 18-APR-2000 (first entry)
XX Maize acetyl CoA carboxylase cDNA.
XX Herbicide resistance; gene modification;
XX maize acetyl CoA carboxylase; ACCase; herbicide; plant oil content;
XX marker-assisted plant selection; groat oil trait;
XX restriction fragment length polymorphism mapping;
XX high-energy animal feed; low-fat human food; ss.
XX Zea mays.
XX Key Location/Qualifiers
XX CDS 37..7014
XX /*tag= a
XX /product= "Maize acetyl CoA carboxylase"
XX 5'UTR 1..36
XX 3'UTR 7014..7470
XX /*tag= c
XX W09967367-A1.
XX 29-DEC-1999.
XX 22-JUN-1999; 99WO-US14022.
XX 22-JUN-1998; 98US-0090240.
XX 02-JUL-1998; 98US-0091640.
XX (MINU ) UNIV MINNESOTA.
XX (USDA ) US DEPT OF AGRICULTURE.
XX (EGLI/) EGLI M A.
XX (GROH/) GROH S.
XX (KIAN/) KIANIAN S F.
XX (PHIL/) PHILLIPS R L.
XX (RINE/) RINES H W.
XX (SOME/) SOMERS D A.
XX Egli MA, Groh S, Kianian SF, Phillips RL, Rines HW, Somers DA;
XX WPI; 2000-147205/13.
XX P-PSDB; Y44687.
XX New DNA encoding acetyl-CoA carboxylase from oats, used to produce
XX transformed plants with herbicide resistance and altered oil content
XX Example 6; Fig 13; 197pp; English.
XX The present cDNA sequence encodes maize acetyl CoA carboxylase (ACCase).
XX ACCase is involved in fatty acid synthesis and is the target of
XX the specified herbicides. Transformation of plants with ACCase imparts
XX resistance to cyclohexanedione and aryloxyphenoxypyranoic acid
XX herbicides and alter the oil content. The ACCase DNA is also used as
XX source of probes and primers for the identification of transgenic
XX plants; in marker-assisted plant selection and for restriction fragment
XX length polymorphism mapping, used for high-energy animal feed and
XX high-fiber, low-fat human food and in genetic dissection of the groat
XX oil trait.
XX Sequence 7470 BP; 2119 A; 1430 C; 1843 G; 2071 T; 7 other;

Query Match 73.6%; Score 16.2; DB 21; Length 7470;
Best Local Similarity 85.7%; Pred. No. 58;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 gttggatcacaaagatttgata 22
Db 3972 gttggatgacaaagatttgata 3992
||||| ||||| |||||
```

•
Search completed: March 27, 2001, 08:23:02
Job time: 5304 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 27, 2001, 08:19:23 ; Search time 132.3 Seconds
(without alignments)
26.799 Million cell updates/sec

Title: US-09-380-826A-7
Perfect score: 22
Sequence: 1 tgttgatcacaaagtattgata 22

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 280836 seqs, 80580151 residues

Total number of hits satisfying chosen parameters: 561672

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents,NA:*

- 1: /cgn2.6/ptodata/2/ina/5A_COMB.seq:*
- 2: /cgn2.6/ptodata/2/ina/5B_COMB.seq:*
- 3: /cgn2.6/ptodata/2/ina/6_COMB.seq:*
- 4: /cgn2.6/ptodata/2/ina/PTUS_COMB.seq:*
- 5: /cgn2.6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	16.2	73.6	2000	1	US-08-014-326-1
2	16.2	73.6	2001	3	US-08-417-089-1
3	16.2	73.6	2001	3	US-08-695-651-1
4	16.2	73.6	4345	2	US-08-244-537-1
5	16.2	73.6	7470	3	US-08-417-089-5
6	16.2	73.6	7470	3	US-08-695-651-5
7	15.6	70.9	296	3	US-08-602-145-14
8	15.6	70.9	1096	1	US-08-684-862-8
9	15.6	70.9	1989	2	US-08-792-055-1
10	15.6	70.9	7745	1	US-08-299-675-1
11	15.6	70.9	7745	1	US-08-485-241-1
12	15.6	70.9	7745	2	US-08-874-162-1
13	15.2	69.1	1016	1	US-08-399-986B-3
14	15.2	69.1	1016	1	US-08-493-754A-3
15	15.2	69.1	2182	1	US-08-399-986B-1
16	15.2	69.1	2182	3	US-08-493-754A-1
17	15.2	69.1	2598	3	US-08-745-892-20
18	15.2	69.1	5420	5	5256642-3
19	15.2	69.1	5420	5	5472939-3
20	15.2	69.1	6951	5	5256642-1
21	15.2	69.1	6951	5	5472939-1
22	14.8	67.3	1855	3	US-08-961-083-71
23	14.8	67.3	3159	1	US-08-119-361-4
24	14.8	67.3	3159	3	US-08-336-308A-3
25	14.8	67.3	3159	3	US-08-822-324-3
26	14.8	67.3	7266	3	US-08-336-308A-9
27	14.8	67.3	7266	3	US-08-822-324-5
28	14.8	67.3	8640	1	US-08-570-311-28

29	14.8	67.3	9919	3	US-08-880-179-1	Sequence 1, Appli
30	14.6	66.4	293	3	US-08-866-340-13	Sequence 13, Appli
31	14.6	66.4	1404	1	US-08-204-656B-1	Sequence 1, Appli
32	14.6	66.4	1404	1	US-08-204-656B-3	Sequence 3, Appli
33	14.6	66.4	1404	1	US-08-204-656B-5	Sequence 5, Appli
34	14.6	66.4	1404	1	US-08-204-656B-7	Sequence 7, Appli
35	14.6	66.4	1404	1	US-08-470-702-1	Sequence 1, Appli
36	14.6	66.4	1404	1	US-08-470-702-2	Sequence 2, Appli
37	14.6	66.4	1404	1	US-08-470-702-3	Sequence 3, Appli
38	14.6	66.4	1404	1	US-08-470-702-4	Sequence 4, Appli
39	14.6	66.4	1404	1	US-08-467-831-1	Sequence 1, Appli
40	14.6	66.4	1404	1	US-08-467-831-2	Sequence 2, Appli
41	14.6	66.4	1404	1	US-08-467-831-3	Sequence 3, Appli
42	14.6	66.4	1404	1	US-08-467-831-4	Sequence 4, Appli
c 43	14.6	66.4	1608	2	US-08-424-224-1	Sequence 1, Appli
c 44	14.6	66.4	1608	4	PCT-US94-02891-68	Sequence 68, Appli
45	14.6	66.4	12047	2	US-09-022-461-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-014-326-1
; Sequence 1, Application US/08014326
; Patent No. 5498544
; GENERAL INFORMATION:
; APPLICANT: Gengenbach, Burle G.
; APPLICANT: Somers, David A.
; APPLICANT: Wyse, Donald L.
; APPLICANT: Gronwald, John W.
; APPLICANT: Egli, Margaret A.
; APPLICANT: Lutz, Shieila M.
; TITLE OF INVENTION: Method and An Acetyl CoA Carboxylase
; TITLE OF INVENTION: Alteration in Oil Content of Plants
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant & Gould
; STREET: 3100 No. 5498544west Center
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PC-DOS/MS-DOS
CURRENT APPLICATION DATA: PC-DOS/MS-DOS
APPLICATION NUMBER: US/08/014,326
FILING DATE: 05-FEB-1993
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/917,462
FILING DATE: 21-JUL-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/538,674
FILING DATE: 18-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/269,584
FILING DATE: 10-NOV-1988
ATTORNEY/AGENT INFORMATION:
NAME: Kowalchuk, Katherine M.
REGISTRATION NUMBER: 36,848
REFERENCE/DOCKET NUMBER: 600.258-US01
TELEPHONE: 612-332-5300
TELEFAX: 612-332-9081
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2000 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

```
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; IMMEDIATE SOURCE:
; CLONE: 2 kb fragment of lambda clone #15-14
US-08-014-326-1

Query Match 73.6%; Score 16.2; DB 1; Length 2000;
Best Local Similarity 85.7%; Pred. No. 20;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 gttggatcacaaagatttgata 22
   ||||| ||||| ||| ||
Db 42 GTTGGATGACAAAGATTGTGA 62

RESULT 2
US-08-417-089-1
; Sequence 1, Application US/08417089
; Patent No. 6069298
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: METHODS AND AN ACETYL COA CARBOXYLASE GENE
; TITLE OF INVENTION: FOR CONFERRING HERBICIDE TOLERANCE AND AN ALTERATION IN
; TITLE OF INVENTION: OIL CONTENT OF PLANTS
; NUMBER OF SEQUENCES: 11
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/417.089
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2001 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-417-089-1

Query Match 73.6%; Score 16.2; DB 3; Length 2001;
Best Local Similarity 85.7%; Pred. No. 20;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 gttggatcacaaagatttgata 22
   ||||| ||||| ||| ||
Db 42 GTTGGATGACAAAGATTGTGA 62

RESULT 3
US-08-695-651-1
; Sequence 1, Application US/08695651
; Patent No. 6146867
; GENERAL INFORMATION:
; APPLICANT: Gengenbach, B. G.
; APPLICANT: Somers, D. A.
; APPLICANT: Wyse, D. L.
; APPLICANT: Gronwald, J. W.
; APPLICANT: Egli, M. A.
; APPLICANT: Lutz, S. M.
; TITLE OF INVENTION: METHOD FOR PRODUCING PLANT ACETYL COA
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schweeman, Lundberg, Woessner & Kluth, P.A.
; STREET: P.O. Box 2938
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
```

```
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/695.651
; FILING DATE: 12-AUG-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/417089
; FILING DATE: 05-APR-1995
; APPLICATION NUMBER: 08/014326
; FILING DATE: 05-FEB-1993
; APPLICATION NUMBER: 07/917462
; FILING DATE: 21-JUL-1992
; APPLICATION NUMBER: 07/538674
; FILING DATE: 18-JUN-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Woessner, Warren D
; REGISTRATION NUMBER: 30,440
; REFERENCE/DOCKET NUMBER: 600.318US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-373-5900
; TELEFAX: 612-339-3061
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2001 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
US-08-695-651-1

Query Match 73.6%; Score 16.2; DB 3; Length 2001;
Best Local Similarity 85.7%; Pred. No. 20;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 gttggatcacaaagatttgata 22
   ||||| ||||| ||| ||
Db 42 GTTGGATGACAAAGATTGTGA 62

RESULT 4
US-08-244-537-1
; Sequence 1, Application US/08244537
; Patent No. 5854420
; GENERAL INFORMATION:
; APPLICANT: ASHTON, ANTHONY R.
; APPLICANT: JENKINS, COLIN L.D.
; APPLICANT: WHITFIELD, PAUL R.
; TITLE OF INVENTION: MAIZE ACETYL COA CARBOXYLASE ENCODING
; TITLE OF INVENTION: DNA CLONES
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: D. C.
; COUNTRY: U.S.A.
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/244.537
; FILING DATE: 18-AUG-1994
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9125330.2
```


;; FILING DATE: 28-NOV-1991
;; PRIORITY APPLICATION DATA:
;; APPLICATION NUMBER: PCT/GB92/02205
;; FILING DATE: 27-NOV-1992
;; ATTORNEY/AGENT INFORMATION:
;; NAME: KOKULIS, PAUL N.
;; REGISTRATION NUMBER: 16,773
;; REFERENCE/DOCKET NUMBER: 203094/SEE 36663/UST
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (202) 861-3000
;; TELEFAX: (202) 822-0944
;; TELEX: 6714627 CUSH
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 4345 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: both
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
US-08-244-537-1

Query Match 73.6%; Score 16.2; DB 2; Length 4345;
Best Local Similarity 85.7%; Pred. No. 22;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 gttggatcacaaagatttgata 22
||||| ||||| ||| ||
Db 902 GTTGGATGACAAAGATTGTGA 922

RESULT 5
US-08-417-089-5
; Sequence 5, Application US/08417089
; Patent No. 6059298
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: METHODS AND AN ACETYL COA CARBOXYLASE GENE
; TITLE OF INVENTION: FOR CONFERRING HERBICIDE TOLERANCE AND AN ALTERATION IN
; TITLE OF INVENTION: OIL CONTENT OF PLANTS
; NUMBER OF SEQUENCES: 11
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/417,089
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7470 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-417-089-5

Query Match 73.6%; Score 16.2; DB 3; Length 7470;
Best Local Similarity 85.7%; Pred. No. 23;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 gttggatcacaaagatttgata 22
||||| ||||| ||||| ||
Db 3972 GTTGGATGACAAAGATTGTGA 3992

RESULT 6
US-08-695-651-5
; Sequence 5, Application US/08695651
; Patent No. 6146867
; GENERAL INFORMATION:
; APPLICANT: Gengenbach, B. G.

;; APPLICANT: Somers, D. A.
;; APPLICANT: Wyse, D. L.
;; APPLICANT: Gronwald, J. W.
;; APPLICANT: Egli, M. A.
;; APPLICANT: Latz, S. M.
;; TITLE OF INVENTION: METHOD FOR PRODUCING PLANT ACETYL COA
;; NUMBER OF SEQUENCES: 11
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Schwegman, Lundberg, Woessner & Kluth, P. A.
;; STREET: P.O. Box 2938
;; CITY: Minneapolis
;; STATE: MN
;; COUNTRY: USA
;; ZIP: 55402
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette
;; COMPUTER: IBM Compatible
;; OPERATING SYSTEM: DOS
;; SOFTWARE: FastSeq Version 2.0
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/695,651
;; FILING DATE: 12-AUG-1996
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/417089
;; FILING DATE: 05-APR-1995
;; APPLICATION NUMBER: 08/014326
;; FILING DATE: 05-FEB-1993
;; APPLICATION NUMBER: 07/917462
;; FILING DATE: 21-JUL-1992
;; APPLICATION NUMBER: 07/538674
;; FILING DATE: 18-JUN-1990
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Woessner, Warren D
;; REGISTRATION NUMBER: 30,440
;; REFERENCE/DOCKET NUMBER: 600.318US3
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 612-373-6900
;; TELEFAX: 612-339-3061
;; TELEX:
;; INFORMATION FOR SEQ ID NO: 5:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 7470 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
US-08-695-651-5

Query Match 73.6%; Score 16.2; DB 3; Length 7470;
Best Local Similarity 85.7%; Pred. No. 23;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 gttggatcacaaagatttgata 22
||||| ||||| ||||| ||
Db 3972 GTTGGATGACAAAGATTGTGA 3992

RESULT 7
US-08-602-145-14/c
; Sequence 14, Application US/08602145
; Patent No. 6025336
; GENERAL INFORMATION:
; APPLICANT: Goltry, Kristin L.
; APPLICANT: Greenberger, Joel S.
; TITLE OF INVENTION: DETERMINING EXPOSURE TO IONIZING RADIATION
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington

```

INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 1096 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
ORIGINAL SOURCE:
ORGANISM: Agkistrodon rhodostoma
FEATURE:
LOCATION: 144 to 841
OTHER INFORMATION: the coding region shown in (2)(ix)(B)
OTHER INFORMATION: codes for the protein of SEQ ID NO: 3
US-08-684-862-8

Query Match 70.9%; Score 15.6; DB 1; Length 1096;
Best Local Similarity 81.8%; Pred. No. 37;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ttttgatcacagatttgata 22
||||| ||||| ||||| ||
Db 112 tctagtatcacagtttgcta 91

RESULT 9
US-08-792-055-1/c
; Sequence 1, Application US/08792055
; Patent No. 5853980
; GENERAL INFORMATION:
; APPLICANT: Rollin, Pierre E.
; APPLICANT: Elliott, Luanne
; APPLICANT: Ksiazek, Thomas G.
; APPLICANT: Nichol, Stuart T.
; APPLICANT: Morzunov, Sergey
; APPLICANT: Ravkov, Eugeny
; TITLE OF INVENTION: The Black Creek Canal Hantavirus and
; TITLE OF INVENTION: Related Methods
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: 127 Peachtree Street, N.E., Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/792.055
; FILING DATE: 03-FEB-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/390,361
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Spratt, Gwendolyn D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 1414.622
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 688-0770
; TELEFAX: (404) 688-9880
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1989 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-792-055-1

```

```
Query Match 70.9%; Score 15.6; DB 2; Length 1989;
Best Local Similarity 81.8%; Pred. No. 39;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ttttgatcacaaagatttgata 22
||||| ||||| ||||| ||
Db 288 TGTGGATCAACAGATTGTA 267

RESULT 10
US-08-299-675-1/c
; Sequence 1, Application US/08299675
; Patent No. 5589618
; GENERAL INFORMATION:
; APPLICANT: Hannah, L. Curtis
; APPLICANT: Giroux, Michael
; TITLE OF INVENTION: Materials and Methods for Increasing
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/299,675
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/299,675
; FILING DATE: 1-SEP-1994
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Pace, Doran R.
; REGISTRATION NUMBER: 38,261
; REFERENCE/DOCKET NUMBER: UF146.C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7745 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; APPLICATION NUMBER: US/08/299,675
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: UF146
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7745 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-299-675-1

Query Match 70.9%; Score 15.6; DB 1; Length 7745;
Best Local Similarity 81.8%; Pred. No. 46;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ttttgatcacaaagatttgata 22
||||| ||||| ||||| ||
Db 3600 TGTGGATCAACAGATGACATA 3579

RESULT 11
US-08-485-241-1/c
; Sequence 1, Application US/08485241
; Patent No. 5650557
; GENERAL INFORMATION:
; APPLICANT: Hannah, L. Curtis
; APPLICANT: Giroux, Michael
; TITLE OF INVENTION: Materials and Methods for Increasing
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/874,162
; FILING DATE: 13-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
```

```
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,241
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/299,675
; FILING DATE: 1-SEP-1994
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Pace, Doran R.
; REGISTRATION NUMBER: 38,261
; REFERENCE/DOCKET NUMBER: UF146.C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7745 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-485-241-1

Query Match 70.9%; Score 15.6; DB 1; Length 7745;
Best Local Similarity 81.8%; Pred. No. 46;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ttttgatcacaaagatttgata 22
||||| ||||| ||||| ||
Db 3600 TGTGGATCAACAGATGACATA 3579

RESULT 12
US-08-874-162-1/c
; Sequence 1, Application US/08874162
; Patent No. 5872216
; GENERAL INFORMATION:
; APPLICANT: Hannah, L. Curtis
; APPLICANT: Giroux, Michael
; TITLE OF INVENTION: Materials and Methods for Increasing
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/874,162
; FILING DATE: 13-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
```

```

* APPLICATION NUMBER: US 08/485,241
  FILING DATE: 7-JUN-1995
  CLASSIFICATION: 435
  PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US 08/299,675
  FILING DATE: 1-SEP-1994
  CLASSIFICATION: 435
  ATTORNEY/AGENT INFORMATION:
  NAME: Pace, Doran R.
  REGISTRATION NUMBER: 38,261
  REFERENCE/DOCKET NUMBER: UF-146C1D1
  TELECOMMUNICATION INFORMATION:
  TELEPHONE: 352-375-8100
  TELEFAX: 352-372-5800
  INFORMATION FOR SEQ ID NO: 1:
  SEQUENCE CHARACTERISTICS:
  LENGTH: 7745 base pairs
  TYPE: nucleic acid
  STRANDEDNESS: single
  TOPOLOGY: linear
  MOLECULE TYPE: DNA (genomic)
US-08-874-162-1

Query Match 70.9%; Score 15.6; DB 2; Length 7745;
Best Local Similarity 81.8%; Pred. No. 46;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ttgtgatcacaaagtattgata 22
||||| ||||| ||||| |||||
Db 3600 TGTGGATAACAAGATGACATA 3579

RESULT 13
US-08-399-986B-3/c
; Sequence 3, Application US/08399986B
; Patent No. 5801041
; GENERAL INFORMATION:
; APPLICANT: Godwin, Andrew K.
; TITLE OF INVENTION: No. 5801041el Gene Associated with Suppression
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/399,986B
; FILING DATE: 06-MAR-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Hagan, Patrick J.
; REGISTRATION NUMBER: 27,643
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1016 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: not relevant
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO

```

```

US-08-399-986B-3

Query Match 69.1%; Score 15.2; DB 1; Length 1016;
Best Local Similarity 85.0%; Pred. No. 57;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 ttggatcacaaagtattgata 22
||||| ||||| ||||| |||||
Db 830 TTGGGTCTCAAGAATTGATA 811

RESULT 14
US-08-493-754A-3/c
; Sequence 3, Application US/08493754A
; Patent No. 5821338
; GENERAL INFORMATION:
; APPLICANT: Godwin, Andrew K.
; TITLE OF INVENTION: No. 5821338el Gene Associated with Suppression
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/493,754A
; FILING DATE: 22-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hagan, Patrick J.
; REGISTRATION NUMBER: 27,643
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1016 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: not relevant
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-493-754A-3

Query Match 69.1%; Score 15.2; DB 1; Length 1016;
Best Local Similarity 85.0%; Pred. No. 57;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 ttggatcacaaagtattgata 22
||||| ||||| ||||| |||||
Db 830 TTGGGTCTCAAGAATTGATA 811

RESULT 15
US-08-399-986B-1/c
; Sequence 1, Application US/08399986B
; Patent No. 5801041
; GENERAL INFORMATION:
; APPLICANT: Godwin, Andrew K.
; TITLE OF INVENTION: No. 5801041el Gene Associated with Suppression
; NUMBER OF SEQUENCES: 35

```

;; - CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
;; STREET: 1601 Market Street
;; CITY: Philadelphia
;; STATE: PA
;; COUNTRY: USA
;; ZIP: 19103-2307
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent In Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/399,986B
;; FILING DATE: 06-MAR-1995
;; CLASSIFICATION: 530
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Hagan, Patrick J.
;; REGISTRATION NUMBER: 27,643
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (215) 563-4100
;; TELEFAX: (215) 563-4044
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 2182 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: not relevant
;; MOLECULE TYPE: DNA (genomic)
;; HYPOTHETICAL: NO
;; ANTI-SENSE: NO
US-08-399-986B-1

Query Match 69.1%; Score 15.2; DB 1; Length 2182;
Best Local Similarity 85.0%; Pred. No. 62;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 ttggatcacaaagatttgata 22
||| || |||| |||||
Db 1996 TTGGGTCTCAGAAATTGATA 1977

Search completed: March 27, 2001, 08:19:25
Job time: 5152 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 27, 2001, 07:38:25 ; Search time 2517.78 Seconds
(without alignments)
61.230 Million cell updates/sec

Title: US-09-380-826A-7
Perfect score: 22
Sequence: 1 tggatgcacaaagatttgata 22

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues

Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: em_estba:*
42: em_estfun:*
43: em_esthum1:*

44: em_esthum2:*
45: em_esthum3:*
46: em_esthum4:*
47: em_esthum5:*
48: em_esthum6:*
49: em_esthum7:*
50: em_esthum8:*
51: em_esthum9:*
52: em_esthum10:*
53: em_esthum11:*
54: em_esthum12:*
55: em_esthum13:*
56: em_esthum14:*
57: em_esthum15:*
58: em_esthum16:*
59: em_esthum17:*
60: em_esthum18:*
61: em_esthum19:*
62: em_esthum20:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estov1:*
68: em_estov2:*
69: em_estpl1:*
70: em_estpl2:*
71: em_estpl3:*
72: em_estpl4:*
73: em_estpl5:*
74: em_estpl6:*
75: em_estro1:*
76: em_estro2:*
77: em_estro3:*
78: em_estro4:*
79: em_estro5:*
80: em_estro6:*
81: em_estro7:*
82: em_estro8:*
83: em_estro9:*
84: em_estro10:*
85: em_estro11:*
86: em_estro12:*
87: em_estro13:*
88: gb_est41:*
89: gb_est42:*
90: gb_est43:*
91: gb_est44:*
92: gb_est45:*
93: gb_est46:*
94: gb_est47:*
95: gb_est48:*
96: gb_est49:*
97: gb_est50:*
98: gb_est51:*
99: gb_est52:*
100: gb_est53:*
101: gb_est54:*
102: gb_est55:*
103: gb_est56:*
104: gb_est57:*
105: gb_est58:*
106: gb_est59:*
107: gb_est60:*
108: gb_est61:*
109: gb_est62:*
110: gb_est63:*
111: gb_est64:*
112: em_esthum21:*
113: em_esthum22:*
114: em_esthum23:*
115: em_estom1:*
116: em_estom2:*

117: em_estp16:*
118: em_estp17:*
119: em_estp18:*
120: em_estp19:*
121: em_estp20:*
122: em_estp21:*
123: em_estp22:*
124: em_estp23:*
125: em_estp24:*
126: em_estp25:*
127: em_estp26:*
128: em_estp27:*
129: em_estp28:*
130: em_estp29:*
131: em_estp30:*
132: em_estp31:*
133: em_estp32:*
134: em_estp33:*
135: em_estp34:*
136: em_estp35:*
137: em_estp36:*
138: em_estp37:*
139: em_estp38:*
140: em_estp39:*
141: em_estp40:*
142: em_estp41:*
143: em_estp42:*
144: em_estp43:*
145: em_estp44:*
146: em_estp45:*
147: em_estp46:*
148: em_estp47:*
149: em_estp48:*
150: em_estp49:*
151: em_estp50:*
152: em_estp51:*
153: em_estp52:*
154: em_estp53:*
155: em_estp54:*
156: em_estp55:*
157: em_estp56:*
158: em_estp57:*
159: em_estp58:*
160: em_estp59:*
161: em_estp60:*
162: em_estp61:*
163: em_estp62:*
164: em_estp63:*
165: em_estp64:*
166: em_estp65:*
167: em_estp66:*
168: em_estp67:*
169: em_estp68:*
170: em_estp69:*
171: em_estp70:*
172: em_estp71:*
173: em_estp72:*
174: em_estp73:*
175: em_estp74:*
176: em_estp75:*
177: em_estp76:*
178: em_estp77:*
179: em_estp78:*
180: em_estp79:*
181: em_estp80:*
182: em_estp81:*
183: em_estp82:*
184: em_estp83:*
185: em_estp84:*
186: em_estp85:*
187: em_estp86:*
188: em_estp87:*
189: em_estp88:*

190: gb_gss25:*
191: gb_gss26:*
192: gb_gss27:*
193: gb_gss28:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	18.8	85.5	361	37	AV532040	AV532040 AV532040
C 2	18.8	85.5	394	147	226559	226559 ATTS1651 Ve
C 3	18.8	85.5	490	14	AB038725	AB038725 AB038725
C 4	18.8	85.5	527	37	AV520789	AV520789 AV520789
C 5	18.8	85.5	530	27	AI995637	AI995637 701676626
C 6	18.8	85.5	608	138	BE977618	BE977618 bs66808.y
C 7	18.4	83.6	934	191	CNS02BWA	AL190387 Tetradon
C 8	17.8	80.9	633	138	BF006254	BF006254 EST434942
C 9	17.8	80.9	633	138	BF006254	BF006254 EST434942
C 10	17.4	79.1	286	126	BB288727	BB288727 BB288727
C 11	17.4	79.1	407	110	BE581720	BE581720 kg51c07.y
C 12	17.4	79.1	529	158	AQ437779	AQ437779 HS_5066.B
C 13	17.4	79.1	674	94	AW695022	AW695022 NF082804S
C 14	17.4	79.1	762	106	BE283043	BE283043 601101323
C 15	17.4	79.1	837	110	BE642809	BE642809 CF127_7B0
C 16	17.4	79.1	1101	190	CNS00D26	AL077427 Drosophila
C 17	17.2	78.2	151	134	BE068225	BE068225 MR4-BF036
C 18	17.2	78.2	425	15	AI049911	AI049911 an30h03.x
C 19	17.2	78.2	431	159	AQ595814	AQ595814 HS_2132.B
C 20	17.2	78.2	432	150	AQ221715	AQ221715 HS_2010.A
C 21	17.2	78.2	437	146	W43212	W43212 22591_Lambd
C 22	17.2	78.2	439	162	AQ796989	AQ796989 nbxb0071H
C 23	17.2	78.2	500	24	AI733664	AI733664 an30h03.x
C 24	17.2	78.2	524	174	B62582	B62582 T22F18TR TA
C 25	17.2	78.2	531	150	AQ223477	AQ223477 HS_2003.B
C 26	17.2	78.2	536	39	AW034253	AW034253 EST277824
C 27	17.2	78.2	546	24	AI779714	AI779714 EST260593
C 28	17.2	78.2	579	174	B67515	B67515 T22M10TR TA
C 29	17.2	78.2	602	24	AI777095	AI777095 EST258060
C 30	17.2	78.2	614	105	BE187570	BE187570 EST336131
C 31	17.2	78.2	619	97	AW963686	AW963686 EST375759
C 32	17.2	78.2	680	174	B57784	B57784 CIT-HSP-201
C 33	17.2	78.2	698	168	A2193715	A2193715 SP_1023.B
C 34	17.2	78.2	904	191	CNS029DN	AL187124 Tetradon
C 35	17.2	78.2	949	192	CNS04HNP	AL291166 Tetradon
C 36	17.3	77.3	373	140	C70416	C70416 C70416 YUJ1
C 37	16.8	76.4	238	34	AV313332	AV313332 AV313332
C 38	16.8	76.4	260	104	BE118999	BE118999 UI-R-CA0-
C 39	16.8	76.4	296	145	T20362	T20362 6c01908-t7
C 40	16.8	76.4	396	97	AW943911	AW943911 LD47517.3
C 41	16.8	76.4	397	91	AW487226	AW487226 81727 MAR
C 42	16.8	76.4	432	157	AQ449764	AQ449764 500004C01
C 43	16.8	76.4	446	24	AI761541	AI761541 w161f07.x
C 44	16.8	76.4	450	138	BF002933	BF002933 7g51c09.x
C 45	16.8	76.4	464	156	AQ332265	AQ332265 HS_5007.A

ALIGNMENTS

RESULT 1
AV532040/c
LOCUS AV532040 361 bp mRNA
DEFINITION Arabidopsis thaliana flower buds Columbia Arabidopsis thaliana cDNA clone FB034f09f 3', mRNA sequence.
ACCESSION AV532040
VERSION AV532040.1 GI:8692323
KEYWORDS EST
SOURCE thale cress.

ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 1 to 361)
AUTHORS Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.
TITLE A large scale analysis of cDNA in Arabidopsis thaliana: Generation of 12,028 non-redundant expressed sequence tags from normalized and size-selected cDNA libraries

JOURNAL DNA Res. 7, 175-180 (2000)
MEDLINE 20363093
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES
source
1..361
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone_lib="FB034f09f"
/clone_lib="Arabidopsis thaliana flower buds Columbia"
/tissue_type="flower buds"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 113 a 73 c 61 g 114 t
ORIGIN

Query Match 85.5%; Score 18.8; DB 37; Length 361;
Best Local Similarity 90.9%; Pred. No. 34;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ttgttgatcacgaatttgata 22
||||| ||||||| |||||||
Db 276 TGTGTTCCACAGATTGACA 255

RESULT 2
LOCUS 226559
DEFINITION ATTS1651 Versailles-VC Arabidopsis thaliana cDNA clone VCVDH08 3' similar to Ribonuclease (RNS2), mRNA sequence.
ACCESSION 226559
VERSION 226559.1 GI:404227
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 1 to 394)
AUTHORS CNRS.
TITLE The Arabidopsis thaliana transcribed genome: the GDR cDNA program Unpublished (1996)
JOURNAL Unpublished (1996)
COMMENT Contact: Desprez T., Anselm J., Chiapello H., Rouze P., Caboche M., Hofte H.
INRA Versailles
Laboratoire de Biologie Cellulaire
Route de Saint-Cyr, 78026 Versailles Cedex, France
Email: thierry@versailles.inra.fr.

FEATURES
source
1..394
/organism="Arabidopsis thaliana"
/strain="ecotype Columbia"
/db_xref="taxon:3702"
/clone_lib="VCVDH08"
/clone_lib="Versailles-VC"
/tissue_type="whole seedlings"
/dev_stage="in vitro-grown etiolated seedlings, 5 days old"
/note="Vector: pBluescript"

BASE COUNT 117 a 80 c 63 g 133 t
ORIGIN

Query Match 85.5%; Score 18.8; DB 14; Length 490;
Best Local Similarity 90.9%; Pred. No. 36;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ttgttgatcacgaatttgata 22
||||| ||||||| |||||||
Db 303 TGTGTTCCACAGATTGACA 282

RESULT 4
LOCUS AV520789/c
DEFINITION AV520789 Arabidopsis thaliana aboveground organs two to six-week old Arabidopsis thaliana cDNA clone APZ32g06f 3', mRNA sequence.
ACCESSION AV520789
VERSION AV520789.1 GI:8680316
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana

ORIGIN

Query Match 85.5%; Score 18.8; DB 147; Length 394;
Best Local Similarity 90.9%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ttgttgatcacgaatttgata 22
||||| ||||||| |||||||
Db 321 TGTGTTCCACAGATTGACA 300

RESULT 3
LOCUS AB038725/c
DEFINITION AB038725 Arabidopsis thaliana Above-ground organ from two to six-week old plants Columbia Arabidopsis thaliana cDNA clone APZ30g03_f 3', mRNA sequence.
ACCESSION AB038725
VERSION AB038725.1 GI:7212552
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 1 to 490)
AUTHORS Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.
TITLE A large scale analysis of cDNA in Arabidopsis thaliana: Generation of 12,028 non-redundant expressed sequence tags from normalized and size-selected cDNA libraries

JOURNAL DNA Res. 7, 175-180 (2000)
MEDLINE 20363093
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES
source
1..490
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone_lib="APZ30g03_f"
/clone_lib="Arabidopsis thaliana Above-ground organ from two to six-week old plants Columbia"
/tissue_type="Above-ground organ from two to six-week old plants"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 147 a 99 c 87 g 157 t
ORIGIN

Query Match 85.5%; Score 18.8; DB 14; Length 490;
Best Local Similarity 90.9%; Pred. No. 36;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ttgttgatcacgaatttgata 22
||||| ||||||| |||||||
Db 303 TGTGTTCCACAGATTGACA 282

RESULT 4
LOCUS AV520789/c
DEFINITION AV520789 Arabidopsis thaliana aboveground organs two to six-week old Arabidopsis thaliana cDNA clone APZ32g06f 3', mRNA sequence.
ACCESSION AV520789
VERSION AV520789.1 GI:8680316
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE

AUTHORS

1 (bases 1 to 527)
Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.

JOURNAL

MEDLINE

COMMENT

The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizuekazusa.or.jp, URL: <http://www.kazusa.or.jp/en/plant/>.

FEATURES

source

1..527
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="AP232q06F"
/clone_lib="Arabidopsis thaliana aboveground organs two to six-week old"
/tissue_type="aboveground organs"
/dev_stage="two to six-week old"
/note="vector: pBluescriptII SK-; site_1: EcoRI; site_2: XhoI"

BASE COUNT 157 a 111 c 101 g 158 t

ORIGIN

Query Match 85.5%; Score 18.8; DB 37; Length 527;
Best Local Similarity 90.9%; Pred. No. 36;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ttttgatcacaaagtattgata 22

||||| |||||||||

Db 247 TGTGTTTCAAGATTGACA 236

RESULT 5

A1995637/c

LOCUS

701676626 A. thaliana, Columbia Col-0, inflorescence-1 Arabidopsis thaliana cDNA clone 701676626, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

EST
thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE

AUTHORS

1 (bases 1 to 530)
Chen, J., Momiyama, M., Chan, E., Mooney, M., Cartoon, B., Gilliland, D., Wang, X., Hillman, J., Guegler, K., Kim, C., Doyle, M., Brzoska, P., Gorgone, G., Burns, D., Griffin, J., Mouanoutoua, M., Nguyen, D., Tan, R., Rose, M., Warren, B., Ton, B., Kastury, K., Borillo, C., Carpio, T., Policky, J., Suzuki, G., Argentine, C., Shah, S., Nobruga, A., Murry, L., Turner, C., Krikorian, S., Elder, L. and Hanson, D.
Arabidopsis thaliana Gene Expression MicroArray
Unpublished (1999)

TITLE

JOURNAL

COMMENT

Contact: David Smoller, Ph.D.
Genome Systems, Inc., a wholly owned subsidiary of Incyte Pharmaceuticals, Inc.
4633 World Parkway Circle, St. Louis, MO 63134, USA

Tel: 877-577-2733

Fax: 314-427-3324

Email: service@genomesystems.com.

Location/Qualifiers

1..530

FEATURES

source

/organism="Arabidopsis thaliana"

/cultivar="Columbia Col-0"

/db_xref="taxon:3702"

/clone="701676626"

/clone_lib="A. thaliana, Columbia Col-0, inflorescence-1"

/tissue_type="inflorescence"

/dev_stage="4 - 7 weeks"

/note="vector: pSPORT; site_1: NotI; site_2: SalI; cDNA library was derived from untreated inflorescence tissue from Arabidopsis thaliana, Columbia Col-0, at 4 - 7 weeks. Plants were grown in 1:1:1 peat moss/vermiculite/perlite soil at 22 deg. C +/- 3 deg. C under constant light, and watered with fertilizer. cDNA synthesis was initiated using a NotI-oligo(dT) primer. Double-stranded cDNA was blunted, ligated to SalI adaptors, digested with NotI, size-selected, and cloned into the NotI and SalI sites of the pSPORT vector."

BASE COUNT 163 a 108 c 100 g 158 t 1 others

ORIGIN

Query Match 85.5%; Score 18.8; DB 27; Length 530;
Best Local Similarity 90.9%; Pred. No. 36;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ttttgatcacaaagtattgata 22

||||| |||||||||

Db 254 TGTGTTTCAAGATTGACA 233

RESULT 6

BE977618/c

LOCUS

BE977618 608 bp mRNA EST 04-OCT-2000
BS666H08.y1 Drosophila melanogaster adult testis library Drosophila melanogaster cDNA clone bs66h08 5', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 608)
Andrews, J., Bouffard, G. and Oliver, B.
Drosophila melanogaster testis expressed sequence tags
Unpublished (1999)

Contact: Brian Oliver

Laboratory of Cellular and Developmental Biology

NIDDK, National Institutes of Health

6 Center Drive MSC 2715, Bldg 6, Rm B1-13, Bethesda, MD 20892 USA

Fax: (301) 496 5239

Email: oliver@helix.nih.gov,

<http://www.niddk.nih.gov/intram/people/boliver.htm>

Tissue isolation and library construction performed at the National Institute of Diabetes and Digestive and Kidney Diseases, NIH (see <http://www.niddk.nih.gov/intram/people/boliver.htm>). DNA sequencing and analyses performed by National Institutes of Health Intramural Sequencing Center (NISC; see <http://www.nisc.nih.gov>).

Plate: 66 row: h column: 08

Seq primer: M13RP1 reverse primer (ABI).

Location/Qualifiers

1..608

FEATURES

source

/organism="Drosophila melanogaster"

/strain="y[+] w[67cl]/Y"

/db_xref="taxon:7227"

/clone="bs66h08"

/clone_lib="Drosophila melanogaster adult testis library"

/sex="male"

/dev_stage="1-5 day adult"

/lab_host="SOLR (Stratagene)"

/note="Organ: testis; Vector: pBlueScript SK (Stratagene); site_1: EcoR I; site_2: Xho I; Testes dissected from 1-5 day adult y[+] w[67cl]/Y males raised at 25oc. RNA isolated using Trizol (Life Technologies) and a single

round of Poly(A)+ selection using Oligotex (Qiagen). cDNA library constructed using Stratagene ZAP-cDNA synthesis kit. Oligo dt-primed, size fractionated ~1-6 kb, and directionally cloned at EcoRI and XhoI in Uni-ZAP XR. Following a single round of amplification pBluescript SK phagemids were mass excised. A distribution channel for clones is being sought, but not currently available. Requests for clones cannot be honored."

BASE COUNT 173 a 146 c 161 g 128 t
ORIGIN

Query Match 83.6%; Score 18.4; DB 138; Length 608;
Best Local Similarity 95.0%; Pred. No. 59;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 ttggatcacaaagtattgata 22
||||| |||||||||

Db 74 TTGGATCCACAGATTGATA 55

RESULT 7
LOCUS CNS02BWA 934 bp DNA GSS 12-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence PUC-ori end of clone 254C22 of library G from Tetraodon nigroviridis, genomic survey sequence.

ACCESSION AL190387
VERSION AL190387.1 GI:7828491
KEYWORDS GSS: genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Eurypterygii; Ctenosquamata; Acanthomorpha; Euacanthomorpha; Holacanthopterygii; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodontidae; Tetraodon.

REFERENCE 1 (bases 1 to 934)
AUTHORS Roest-Crolius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.

TITLE Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 934)
AUTHORS Roest-Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.

TITLE Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence

JOURNAL Unpublished
REFERENCE 3 (bases 1 to 934)
AUTHORS Genoscope.

TITLE Direct Submission

JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/tetraodon>.

FEATURES
Location/Qualifiers
1..934
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="254C22"
/clone_lib="G"
/note="Genoscope sequence ID : C0AG254BB11SP1-end ; PUC-ori"

BASE COUNT 282 a 175 c 207 g 259 t 11 others
ORIGIN

Query Match 83.6%; Score 18.4; DB 191; Length 934;
Best Local Similarity 95.0%; Pred. No. 62;

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 ttggatcacaaagtattgata 22
||||| |||||||||

Db 840 TTGGAACACAGATTGATA 821

RESULT 8
LOCUS BF006444/c 600 bp mRNA EST 06-OCT-2000
DEFINITION EST434942 DSLC Medicago truncatula cDNA clone pDSL-41J15, mRNA sequence.

ACCESSION BF006444
VERSION BF006444.1 GI:10706719
KEYWORDS EST.
SOURCE barrel medic.
ORGANISM Medicago truncatula

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago.

REFERENCE 1 (bases 1 to 600)
AUTHORS Fedorova,M., Pierson,B.L., Samac,D.A., Gantt,J.S., Vance,C.P., Gonzales,M.B. and Ellis,L.

TITLE ESTs from Medicago truncatula leaves and cotyledons
JOURNAL Unpublished (2000)
COMMENT Contact: Deborah A. Samac
Department of Plant Pathology
University of Minnesota
495 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108, USA
Tel: 612 625 1243
Fax: 651 649 5058
Email: debys@puccini.crl.umn.edu

University of Minnesota name: M275482e TIGR sequence name: MTLBA567K More information is available at: <http://chrysis.tamu.edu/medicago>
Seq primer: SKmod (CTA gaa CTA gtg gat CC).
Location/Qualifiers
1..600
/organism="Medicago truncatula"
/cultivar="genotype AL7"
/db_xref="taxon:3880"
/clone="pDSL-41J15"
/clone_lib="DSL-41J15"
/tissue_type="leaves and cotyledons"
/dev_stage="mixture of cotyledons from five days old plants and leaves obtained from two weeks old plants"
/lab_host="E. coli strain SOLR"
/note="Vector: pBluescript SK +/-; Site 1: EcoRI; Site 2: XhoI; cDNA was prepared from polyA+ enriched RNA from the mixture of cotyledons of five days old plants and leaves of two weeks old plants. The cDNA was directionally ligated into the Uni-ZAP XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-ZAP phage using Ex-Assist helper phage and propagated in SOLR cells."

BASE COUNT 127 a 120 c 132 g 221 t
ORIGIN

Query Match 80.9%; Score 17.8; DB 138; Length 600;
Best Local Similarity 90.5%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 tggatcacaaagtattgat 21
||||| |||||||||

Db 90 TGTGATCAGATTAGAT 70

RESULT 9
LOCUS BF006254/c 633 bp mRNA EST 06-OCT-2000
DEFINITION EST434752 DSLC Medicago truncatula cDNA clone pDSL-40G21, mRNA sequence.

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago.

REFERENCE 1 (bases 1 to 633)
AUTHORS Fedorova,M., Pierson,B.L., Samac,D.A., Gantt,J.S., Vance,C.P., Gonzales,M.B. and Ellis,L.

TITLE ESTs from Medicago truncatula leaves and cotyledons
JOURNAL Unpublished (2000)
COMMENT Contact: Deborah A. Samac
Department of Plant Pathology
University of Minnesota
495 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108, USA
Tel: 612 625 1243
Fax: 651 649 5058
Email: debys@puccini.crl.umn.edu

University of Minnesota name: M275482e TIGR sequence name: MTLBA567K More information is available at: <http://chrysis.tamu.edu/medicago>
Seq primer: SKmod (CTA gaa CTA gtg gat CC).
Location/Qualifiers
1..600
/organism="Medicago truncatula"
/cultivar="genotype AL7"
/db_xref="taxon:3880"
/clone="pDSL-41J15"
/clone_lib="DSL-41J15"
/tissue_type="leaves and cotyledons"
/dev_stage="mixture of cotyledons from five days old plants and leaves obtained from two weeks old plants"
/lab_host="E. coli strain SOLR"
/note="Vector: pBluescript SK +/-; Site 1: EcoRI; Site 2: XhoI; cDNA was prepared from polyA+ enriched RNA from the mixture of cotyledons of five days old plants and leaves of two weeks old plants. The cDNA was directionally ligated into the Uni-ZAP XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-ZAP phage using Ex-Assist helper phage and propagated in SOLR cells."

BASE COUNT 127 a 120 c 132 g 221 t
ORIGIN

Query Match 80.9%; Score 17.8; DB 138; Length 600;
Best Local Similarity 90.5%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 tggatcacaaagtattgat 21
||||| |||||||||

Db 90 TGTGATCAGATTAGAT 70

Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel: 580 221 7302
Fax: 580 221 7380
Email: radixon@noble.org

Insert Length: 674 Std Error: 0.00
Plate: 082 row: E column: 04
Seq primer: TCACACAGGAACACGCTATGAC.

FEATURES

source

Location/Qualifiers

1..674

/organism="Medicago truncatula"

/db_xref="taxon:3880"

/clone="NF082E04St"

/clone_lib="Developing stem"

/tissue_type="stem"

/dev_stages="pooled developmental"

/note="vector: Lambda Zap; Contains a mixture of

internodal stem segments"

189 a 149 c 133 g 203 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 79.1%; Score 17.4; DB 94; Length 674;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 ttggatcacagaatttg 21

Db 256 TTGGATCACTAGATTGAT 238

RESULT 14

BE283043

LOCUS

DEFINITION 601101323F1 NCL_CGAP_Lu29 Mus musculus cDNA clone IMAGE:3493829 5',

13-JUL-2000

EST

Accession

Version

Keywords

Source

Organism

Reference

Authors

Title

Journal

Comment

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1350

Email: Robert.Strausberg@nih.gov

Tissue Procurement: Gilbert Smith, Ph.D.

cDNA Library preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM8541 row: n column: 06

High quality sequence stop: 661.

FEATURES

source

1..762

/organism="Mus musculus"

/strain="C57BL/6J (f1er1)"

/db_xref="taxon:10090"

/clone="IMAGE:3493829"

/clone_lib="NCL_CGAP_Lu29"

/tissue_type="spontaneous tumor, metastatic to mammary.

stem cell origin."

/lab_host="DH10B"

/note="Organ: lung; Vector: pCMV-SPORT6; Site: 1: SalI;

Site: 2: NotI; Cloned unidirectionally. Primer: Oligo dT.

Library constructed by Life Technologies. Investigator

BASE COUNT 193 a 172 c 207 g 190 t

ORIGIN

providing samples: Gilbert Smith, NIH

Query Match 79.1%; Score 17.4; DB 106; Length 762;

Best Local Similarity 94.7%; Pred. No. 1.9e+02;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ttttgatcacagaatttg 19

Db 692 TTTTGGATCACAGACTTG 710

RESULT 15

BE642809

LOCUS

DEFINITION BE642809 837 bp mRNA EST 01-SEP-2000

CDNA clone Cri2_7_B04 5', mRNA sequence.

Accession

Version

Keywords

Source

Organism

Reference

Authors

Title

Journal

Comment

Contact: Roux SJ

Section of Molecular Cell and Developmental Biology

University of Texas

Biology Building, Room 16, Austin, TX 78712, USA

Tel: 512 471 4238

Fax: 512 232 3402

Email: sroux@uts.cc.utexas.edu

Plate: Cri2_7 row: B column: 04

Seq primer: SP6.

Location/Qualifiers

1..837

/organism="Ceratopteris richardii"

/cultivar="Brogn"

/db_xref="taxon:49495"

/clone="Cri2_7_B04"

/clone_lib="Ceratopteris Spore Library"

/tissue_type="Gametophyte"

/cell_type="Spore"

/dev_stages="20 hours after germination initiation"

/note="Vector: pCMVSPORT6; EST sequence from cDNA library.

CDNA library constructed from mRNA isolated from C.

richardii spores that had developed for 20 hours after

their germination had been initiated by white light."

233 a 158 c 214 g 232 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

Conservative

Mismatches

Indels

Gaps

Qy 1 ttttgatcacagaatttg 19

Db 332 TGTAGGATCACAGACTTG 350

Search completed: March 27, 2001, 07:38:26

Job time: 4569 sec

THIS PAGE BLANK (USPTO)